GENOME-WIDE PROFILING OF H1 LINKER HISTONE VARIANTS IN MOUSE EMBRYONIC STEM CELLS

A Dissertation
Presented to
The Academic Faculty

By

Kaixiang Cao

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy in Biology

Georgia Institute of Technology

May, 2014

GENOME-WIDE PROFILING OF H1 LINKER HISTONE VARIANTS IN MOUSE EMBRYONIC STEM CELLS

Approved by:

Dr. Yuhong Fan, Advisor
School of Biology
Georgia Institute of Technology

Dr. Eva K. Lee
School of Industrial & Systems Engineering
Georgia Institute of Technology

Dr. Alfred H. Merrill
School of Biology
Georgia Institute of Technology

Dr. Ingeborg Schmidt-Krey
School of Biology
Georgia Institute of Technology

Dr. Francesca Storici
School of Biology
Georgia Institute of Technology

Date Approved: December 5, 2013

To my parents.

ACKNOWLEDGEMENTS

First, I would like to thank my thesis advisor, Dr. Yuhong Fan, for her guidance and support throughout these years. All my work would not be possible without her accepting me to her lab and teaching me both experimental skills and critical thinking. Her enthusiasm and patience on solving scientific problems constantly encourage me to pursue my career as a scientist.

I want to express my gratitude to my committee members: Dr. Eva Lee, Dr. Alfred Merrill, Dr. Ingeborg Schmidt-Krey, and Dr. Francesca Storici. Their constructive suggestions and advice have been valuable for my studies.

I want to thank Nathalie Lailler and Dr. Eric Bouhassira at Albert Einstein College of Medicine for alignment of the sequencing reads and providing the analysis tools for this project. I thank Karan Uppal, Drs. Eva Lee and Hongwei Wu at Georgia Tech, and Dr. Christoph Bock at Medical University of Vienna for their expertise and help in bioinformatics analysis.

I am grateful to all past and present Fan lab members, especially Yunzhe Zhang, Magdalena Medrzycki, Chenyi Pan, and Po-Yi Ho, for being my long-term friends, offering immense help and support in my research and life all these years. I have benefited from a collegial atmosphere in the lab which fosters scientific discoveries. I thank Ashwath Kumar, Xiao Dong, Guy P. Cooper Jr. for helping with bioinformatics analysis; Dr. Zheng Liu and Yunzhe Zhang for FISH analysis and generation of $H1d^{FLAG/FLAG}$ mouse models; and Magdalena Medrzycki for DNA methylation analysis. Special thanks to Yunzhe for her incredible patience in training me experimental

techniques in stem cell biology. Her superb experimental skills and amiable personality have benefited me greatly.

Last but not least, I am forever indebted to my parents for their trust and love. Their love is my biggest motivation and it makes me believe that all the hard work will pay off.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

A_{214}	absorbence at 214 nm
ADP	adenosine diphosphate
Alu	Arthrobacter luteus transposable elements
AT%	adenine-thymine content
bp	base pair
BRG1	Brahma protein-like 1
cas	CRISPR-associated
cDNA	complementary DNA
CEAS	cis-regulatory element annotation system
ce^{het}	$H1c^{+/-}H1d^{+/+}H1e^{+/-}$
ce^{KO}	H1c/H1e double null
CHD1	chromodomain helicase DNA binding protein 1
CHD8	chromodomain helicase DNA binding protein 8
ChIP	chromatin immunoprecipitation
ChIP-chip	chromatin immunoprecipitation coupled with microarray analysis
ChIP-seq	chromatin immunoprecipitaition coupled with high-throughput sequencing
CRISPR	clustered regularly interspaced short palindromic repeat
DamID	DNA adenine methyltransferase identification
DAPI	4',6-diamidino-2-phenylindole
DMEM	Dulbecco's modified Eagle's medium
DNA	deoxyribonucleic acid
DNase I	deoxyribonuclease I
DNMT	DNA methyltransferase
DNMT1	DNA methyltransferase 1
DNMT3A	DNA methyltransferase 3 alpha
DNMT3B	DNA methyltransferase 3 beta
DOT1L	DOT1 (disruptor of telomeric silencing)-like histone H3K79 methyltransferase
EB	embryoid body or ethidium bromide
EMSA	electrophoretic mobility shift assay
ENCODE	the encyclopedia of DNA elements
ERV	endogenous retrovirus
ESC	embryonic stem cell
FBS	fetal bovine serum
FDR	false discovery rate
$fH1^0$	FLAG-$H1^0$ over-expressing ESCs
FISH	fluorescence in situ hybridization
FITC	fluorescein isothiocyanate
FLAG-H1d	FLAG-tagged H1d
FRAP	fluorescence recovery after photobleaching
GAPDH	glyceraldehyde 3-phosphate dehydrogenase

GC%	guanine-cytosine content
GFP	green fluorescent protein
H1/nuc	The ratio of H1 variants to nucleosome
$H1c^{Myc}$	$H1c^{+/Myc}H1d^{+/-}H1e^{+/-}$ ESCs
$H1d^{FLAG}$	$H1c^{+/-}H1d^{+/FLAG}H1e^{+/-}$ ESCs
$H1d^{FLAG/FLAG}$	$H1c^{-/-}H1d^{FLAG/FLAG}H1e^{-/-}$
H1d-trans	$H1c^{+/-}H1d^{FLAG/-}H1e^{+/-}$ ESCs
H3K27me3	histone H3 lysine 27 tri-methylation
H3K4me3	histone H3 lysine 4 tri-methylation
H3K9me3	histone H3 lysine 9 tri-methylation
H4K12Ac	histone H4 lysine 12 acetylation
H4K20me3	histone H4 lysine 20 tri-methylation
HAT	histone acetyltransferase
HDAC	histone deacetylase
Hox	homeobox
HP1	heterochromatin protein 1
HPLC	High-performance liquid chromatography
HPRT	hypoxanthine-guanine phosphoribosyltransferase
IACUC	the Institutional Animal Care and Use Committee
IAP	intracisternal A-particle
IgG	immunoglobulin G
IN	input
IP	immunoprecipitation
iPSC	induced pluripotent stem cell
Kb	kilo base pairs
LIF	leukemia inhibitory factor
LINE	long interspersed elements
LTR	long terminal repeat
Mb	mega base pairs
MBD2	Methyl-CpG-binding domain protein 2
MBD3	Methyl-CpG-binding domain protein 3
MeCP2	methyl CpG binding protein 2
mESC	mouse embryonic stem cell
MEF	mouse embryonic fibroblast
MMTV	mouse mammary tumor virus
MNase	micrococcal nuclease
Msx1	Msh homeobox 1
Myc	myelocytomatosis oncogene
Myc-H1c	Myc tagged H1c
MyoD	Myogenic differentiation 1
Nanog	nanog homeobox
NRL	nucleosome repeat length
NuRD	nucleosome remodeling and deacetylation
Oct4	octamer-binding transcription factor 4
p53	tumor protein 53
PCR	polymerase chain reaction

PMSF	phenylmethylsulfonyl fluoride
PRC2	polycomb repressive complex 2
qChIP	quantitative chromatin immunoprecipitation
qRT-PCR	quantitative reverse transcription-PCR
rDNA	ribosomal DNA
RES	"rescue" cells
Rhox	reproductive homeobox
RNA	ribonucleic acid
RNase	ribonuclease
RP-HPLC	reversed phase HPLC
RT	reverse transcription
S.D.	standard deviation
S.E.M	standard error of the mean
SAR	scaffold associated regions
SICER	spatial clustering approach for the identification of ChIP-enriched regions
SINE	short interspersed elements
siRNA	small interference RNA
SirT1	sirtuin (silent mating type information regulation 2 homolog) 1
Sox2	SRY (sex determining region Y)-box 2
Su(var)3-9	suppressor of variegation 3-9
Suv39h1	suppressor of variegation 3-9 homolog 1
Suv39h2	suppressor of variegation 3-9 homolog 2
SV40	simian virus 40
TALENs	transcription activator-like effector nucleases
TKO	triple knockout
TSS	transcription start site
TTS	transcription termination site
UCSC	University of California, Santa Cruz
UHRF1	ubiquitin-like, containing PHD and RING finger domains, 1
UTR	untranslated region
WT	wildtype
ZFN	zinc-finger nucleases

SUMMARY

H1 linker histone facilitates the formation of higher order chromatin structure and is essential for mammalian development. Mice have 11 H1 variants which are differentially regulated and conserved in human. Previous research indicates that H1 regulates the expression of specific genes in mouse embryonic stem cells (ESCs). However, whether individual variants have distinct functions and how H1 participates in gene regulation remain elusive. An investigation of the precise localization of individual H1 variants *in vivo* would facilitate the elucidation of mechanisms underlying chromatin compaction regulated gene expression, while it has been extremely difficult due to the lacking of specific antibodies toward H1 variants. In this dissertation, I have generated a knock-in system in ESCs and shown that the N-terminally tagged H1 proteins are functionally interchangeable to their endogenous counterparts *in vivo*. H1d and H1c are depleted from GC- and gene-rich regions and active promoters, inversely correlated with H3K4me3, but positively correlated with H3K9me3 and associated with characteristic sequence features. Surprisingly, both H1d and H1c are significantly enriched at major satellites, which display increased nucleosome spacing compared with bulk chromatin. While also depleted at active promoters and enriched at major satellites, overexpressed $H1^0$ displays differential binding patterns in specific repetitive sequences compared with H1d and H1c. Depletion of H1c, H1d, and H1e causes pericentric chromocenter clustering and de-repression of major satellites. Collectively, these results integrate the localization of an understudied type of chromatin proteins, namely the H1 variants, into

the epigenome map of mouse ESCs, and demonstrate significant changes at pericentric heterochromatin upon depletion of this epigenetic mark.

CHAPTER 1

INTRODUCTION

1.1 H1 Linker Histone and Its Variants

In eukaryotic cells, nuclear DNA is packaged into chromatin with the facilitation of histone proteins (Wolffe 1998). The nucleosome, the building block of chromatin, consists of an octamer core of four core histones (H2A, H2B, H3 and H4) wrapped by 147 bp of DNA (Olins and Olins 1974; Oudet et al. 1975; Wolffe 1998). The 10-nm "beads on a string" structure represents the first level of DNA compaction, in which DNA is wrapped around repeating units of nucleosomes.((Olins and Olins 1974) and Figure 1.1). H1 linker histone, the fifth histone, binds to DNA entering and exiting nucleosome core particles and the linker DNA, facilitating the folding of chromatin into higher order structure such as the 30-nm fiber ((Noll and Kornberg 1977; Thoma et al. 1979; Allan et al. 1980) and Figure 1.1). H1 is highly basic and has a three-domain structure (Chapman et al. 1976): the short N-terminal tail, the central globular domain and the highly lysine rich C-terminal tail. Both the globular and the C- terminal domains have been shown to be critical for the binding of H1 to chromatin (Hendzel et al. 2004; Brown et al. 2006; Stasevich et al. 2010; Syed et al. 2010). However, the exact position of H1 at the nucleosome and the genome-wide binding profiles of H1 remain largely unknown.

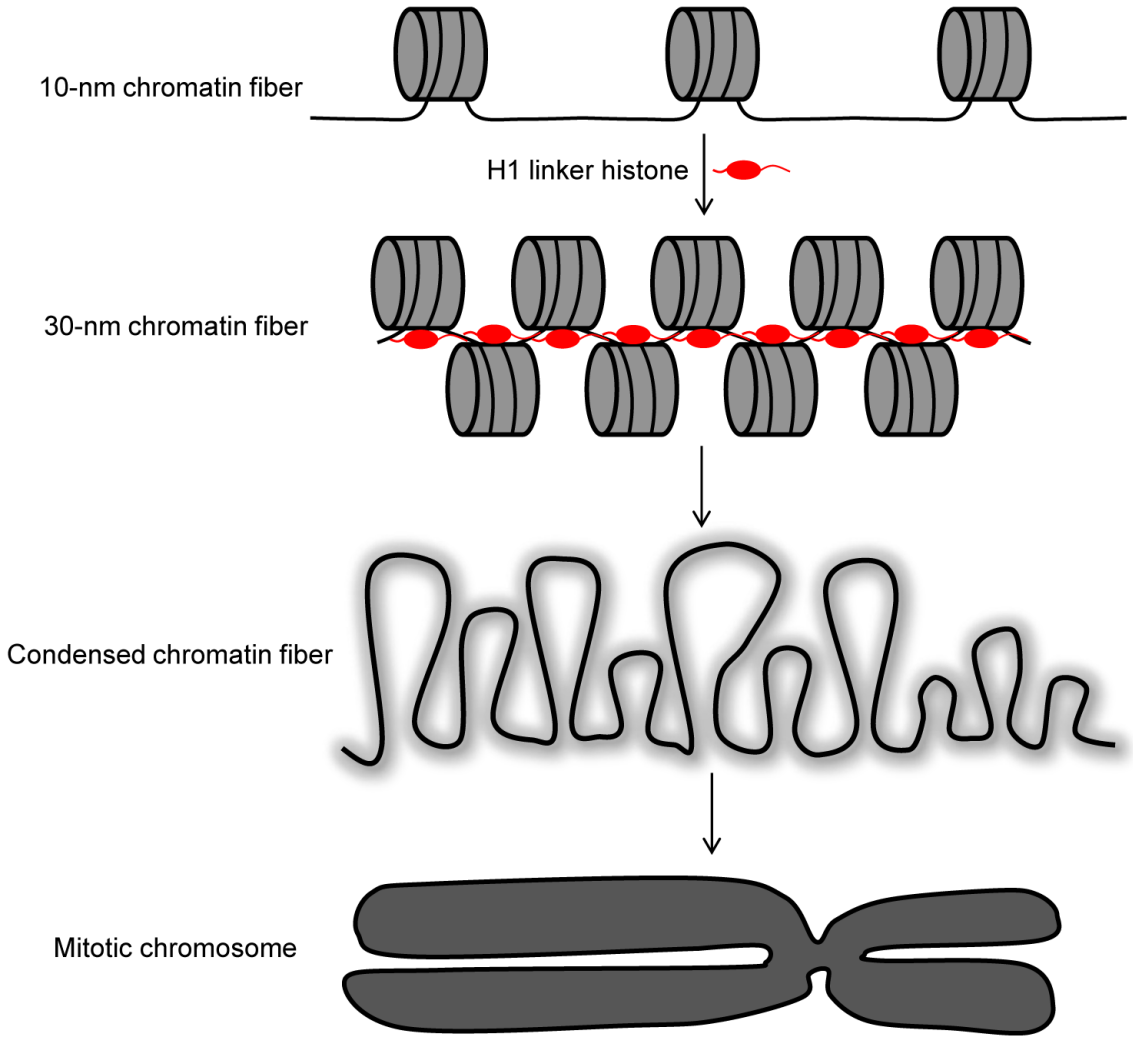

Figure 1.1 The process of DNA packaging into a mitotic chromosome.
H1 facilitates the formation of 30-nm fiber and higher order chromatin structure.

H1 linker histone is the most divergent protein family in histones. To date, there have been 11 different H1 variants characterized in mammals. These variants are conserved from mouse to human and can be categorized into somatic H1s (H1a, H1b, H1c, H1d, H1e, $H1^0$, and H1x) (Albig et al. 1991; Albig et al. 1993; Albig et al. 1997; Happel et al. 2005; Happel et al. 2009) and germ cell specific H1s (H1t, H1T2, HILS1 and H1oo) (Drabent et al. 1991; Drabent et al. 1993; Tanaka et al. 2001; Yan et al. 2003; Martianov et al. 2005). Among the somatic H1 variants, H1a-e are mainly synthesized in the S phase of the cell cycle, while the expression of $H1^0$ and the latest characterized H1 variant, H1x, are cell cycle independent. Protein sequence alignment of mouse H1 variants indicates that the N- and C- terminal domains of somatic H1 variants are more divergent than the highly conserved globular domains (Figure 1.2).

The existence of multiple variants poses a challenge for the functional study of H1 linker histone in mammals. Individual deletion of most H1 variants does not lead to observable phenotype in mouse (Sirotkin et al. 1995; Drabent et al. 2000; Lin et al. 2000; Rabini et al. 2000; Fan et al. 2001), probably due to compensatory effects of the remaining variants. Nevertheless, sequential disruption of three somatic H1 variants (H1c, H1d, and H1e) causes embryonic lethality in mouse (Fan et al. 2003), demonstrating that H1 is essential for mammalian development.

Individual H1 variants have different chromatin binding affinities and residence time (Th'ng et al. 2005; Clausell et al. 2009), probably due to variations in nucleosome binding surface and C-terminal domain (Hendzel et al. 2004; Brown et al. 2006; George et al. 2010; Stasevich et al. 2010; Vyas and Brown 2012). Furthermore, H1 variants are

differentially regulated during development (Lennox and Cohen 1983; Wang et al. 1997), and show distinct distribution in nucleus (Th'ng et al. 2005). These results suggest that specific H1 variants have unique functions. Genome-wide mapping of these variants could reveal their specific binding patterns *in vivo* and shed lights on the roles of these variants in gene regulation.

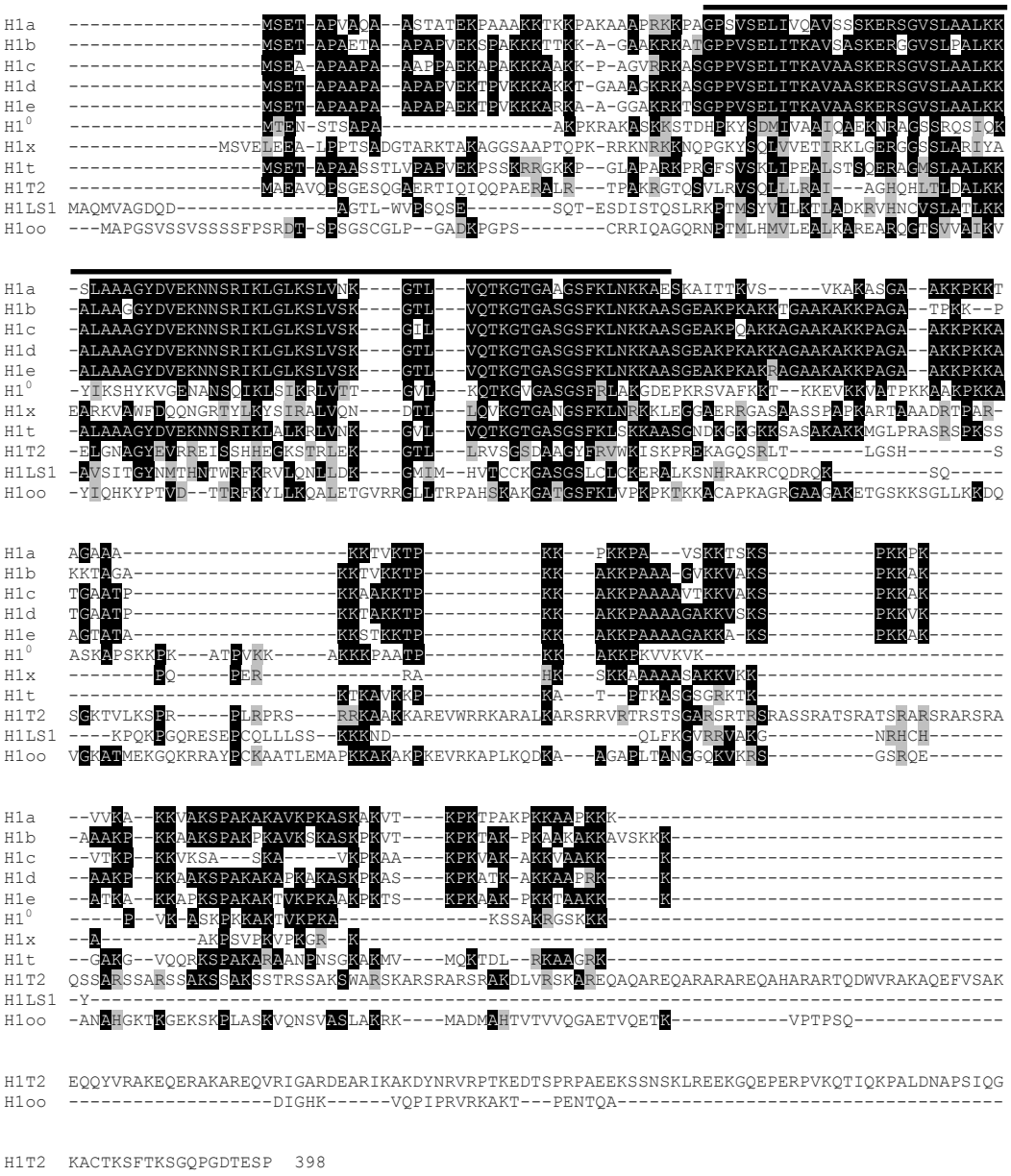

Figure 1.2 Amino acid sequence alignment of mouse H1 variants.
Globular domains are marked by a black line. Identical amino acids are highlighted in black, while similar residues are highlighted by grey.

1.2 Chromatin Binding Features of H1 Linker Histone

H1 is one of the major structural components of chromatin that stabilizes higher order chromatin folding (Thoma and Koller 1977; Allan et al. 1980). Interestingly, recent *in vivo* studies using fluorescence recovery after photobleaching (FRAP) technique and overexpressed GFP-tagged H1 indicates that, rather than just being a static structural component, H1 binds dynamically to chromatin with a rapid exchange rate (Lever et al. 2000; Misteli et al. 2000). Later studies with similar strategies have shown that the chromatin binding affinity of H1 variants is different *in vivo* (Hendzel et al. 2004; Th'ng et al. 2005; Stasevich et al. 2010). However, FRAP analysis does not have the resolution to identify the precise genomic localizations of histone H1.

Chromatin immunoprecipitation (ChIP) is an important technique to study the interaction of a protein and genomic DNA *in vivo* (Hecht et al. 1996) (Figure 1.3). ChIP enriched DNA could be analyzed by quantitative PCR (qChIP) to detect the binding levels of proteins at regions of interest. Moreover, ChIP could be coupled with methods such as tiling array (ChIP-chip) and deep sequencing (ChIP-seq) to analyze the genome-wide localization of specific proteins. Using an antibody recognizing all H1 variants and tiling array covering promoter regions, a recent study shows that H1 is depleted from active promoters in human breast cancer cells by ChIP-chip (Krishnakumar et al. 2008). However, this study could not differentiate individual H1 variants because the antibody used is not variant specific and the tiling array for ~1500 promoters only covers 2% of the human genome. More comprehensive studies on genome-wide binding of specific

H1 variants would elucidate the mechanisms of epigenetic gene regulation mediated by linker histone H1 and higher order chromatin structure.

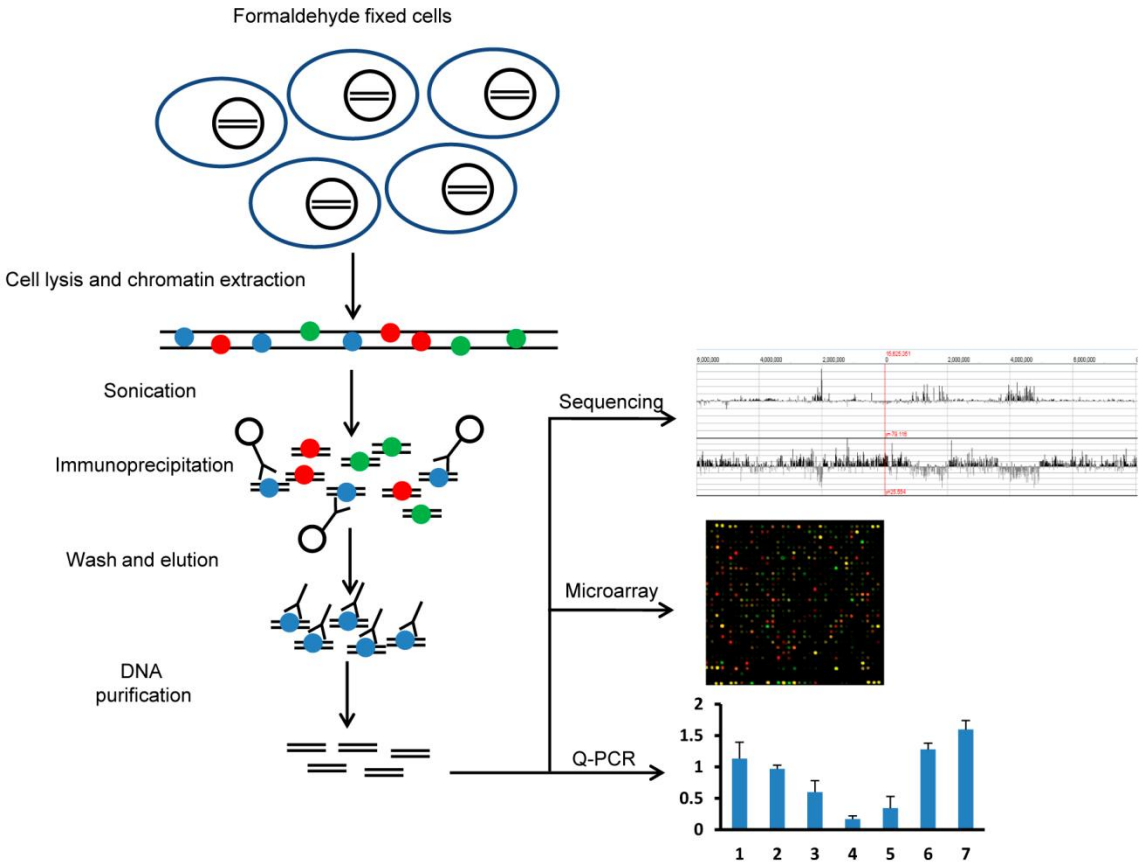

Figure 1.3 Schematic representation of chromatin immunoprecipitation (ChIP) assays.
ChIP enriched DNA could be analyzed by techniques such as qPCR (qChIP), microarray (ChIP-chip), and deep sequencing (ChIP-seq).

1.3 H1 Linker Histone and Gene Regulation

H1 has long been thought to be a general transcription repressor as *in vitro* analyses indicate that H1 counteracts the activation effects of transcription factors (Shimamura et al. 1989; Croston et al. 1991; Laybourn and Kadonaga 1991). However, recent studies suggest that H1 participates in gene regulation processes in a specific manner *in vivo*.

H1 inactivation in lower eukaryotes leads to specific gene expression changes without perturbing global transcription (Shen and Gorovsky 1996; Hellauer et al. 2001; Jedrusik and Schulze 2001; Ni et al. 2006; Lu et al. 2009b; Lu et al. 2013). Interestingly, knockdown of H1.1 but not other variants leads to gene activation limited to germline in *C. elegans* (Jedrusik and Schulze 2001). In higher eukaryotes, specific gene expression changes have been observed in *Xenopus Laevis* (Bouvet et al. 1994; Kandolf 1994; Steinbach et al. 1997) and chicken B-lymphocytes following H1 inactivation (Takami and Nakayama 1997; Takami et al. 2000; Hashimoto et al. 2010). In mouse, transcription alteration has not been reported in most H1 single knockout studies (Sirotkin et al. 1995; Lin et al. 2000; Rabini et al. 2000; Fan et al. 2001; Fantz et al. 2001; Martianov et al. 2005; Tanaka et al. 2005). Nevertheless, by crossing single H1 depleted mice with mice carrying transgenes, individual variants are shown to have distinct effects in attenuating gene silencing induced by position effects (Alami et al. 2003). Double knockout of H1a and H1t in male germ cells (Lin et al. 2004) and triple knockout (TKO) of H1c, H1d and H1e in embryonic stem cells (ESCs) (Fan et al. 2005) cause up and down regulations of specific gene expression. Interestingly, multiple homeobox (Hox) genes are significantly

down-regulated in H1 TKO mouse embryos (Zhang et al. 2012b), providing a possible mechanistic explanation for the developmental defects of such embryos (Fan et al. 2003). In human breast cancer cells, knockdown of individual H1 variants by siRNA leads to expression alterations of a limited number of genes. Specifically, H1c knockdown causes repression of many cell cycle-related genes (Sancho et al. 2008).

$H1^0$ and H1c overexpression in mouse fibroblasts leads to both positively and negatively perturbed gene expressions, some of which are variant specific (Bhan et al. 2008). Similar results have been reported in *Drosophila* after overexpressing H1 (Ni et al. 2006). MMTV promoter has been shown to be particularly sensitive to H1 dosage. Overexpression of $H1^0$ and H1c significantly increases the activity of stably integrated mouse mammary tumor virus (MMTV) promoter, while such elevated activity was not observed in the transiently transfected reporter (Gunjan and Brown 1999). In *Xenopus* oocytes, sub-saturating level of H1 has been shown to activate MMTV promoter (Belikov et al. 2007).

Recent studies also indicate that H1 regulates gene expression through protein-protein interactions. It was reported that H1b and Msx1 interacts and cooperates to inhibit MyoD and myogenic differentiation in mouse myoblasts (Lee et al. 2004). SirT1, a histone deacetylase, has been shown to recruit H1 to a transgene promoter by directly interacting with H1e, which leads to transcription repression (Vaquero et al. 2004). Recently it was shown that CHD8 requires the recruitment of H1 to inhibit p53 and β-catenin dependent transactivation by forming a trimeric complex with p53 and β-catenin, respectively (Nishiyama et al. 2009; Nishiyama et al. 2012). In *Drosophila*, the interaction between H1 and the H3 lysine 9 (H3K9) methyltransferase Su(var)3-9 has

been suggested to suppress the transcription of transposable elements (Lu et al. 2013). The interaction of DNA methyltransferases (DNMTs) with H1 was identified as a potential mechanism of H1's regulation on several imprinted genes in ESCs ((Yang et al. 2013); Cao K and Fan Y, unpublished observation).

In summary, H1 specifically regulates gene expression *in vivo* rather than repressing transcription globally. It is possible that H1 modulates local transcription level through recruitment of interacting proteins.

1.4 H1 Linker Histone and Core Histone Modifications

Core histones are subject to multiple types of post-translational modifications, such as acetylation, methylation, ubiquitylation, sumoylation, phoshphorylation, ADP ribosylation, deimination and proline isomerization (Kouzarides 2007). Lysine acetylation and methylation have been the most extensively studied modifications. Triple H1 depletion in mouse ESC leads to a decrease of two core histone modifications, H4K12Ac and H3K27me3, in bulk chromatin (Fan et al. 2005), indicating that H1 may regulate the levels of core histone modifications.

Histone lysine acetylation is catalyzed by histone acetyltransferases (HATs) and erased by histone deacetylases (HDACs) (Shahbazian and Grunstein 2007). Hyperacetylation is associated with gene activation while hypoacetylation is correlated with repression of transcription. Whole-genome mapping of histone acetylation marks demonstrates that acetylation marks are generally positively correlated with gene expression (Wang et al. 2008). On the other hand, histone lysine methylation marks can

be grouped into either transcription activation marks, such as H3K4me3, or repression marks, such as H3K9me3, H3K27me3, and H4K20me3. For instance, active promoter regions are often marked by H3K4me3 (Guenther et al. 2007; Mikkelsen et al. 2007), whereas H3K9me3 and H3K27me3 are enriched at pericentric and facultative heterochromatin, respectively (Peters et al. 2001; Lehnertz et al. 2003; Trojer and Reinberg 2007). Both H3K9me3 and H4K20me3 are enriched at repetitive sequences (Barski et al. 2007; Mikkelsen et al. 2007). H3K27me3 colocalizes with H3K4me3, forming the "bivalent domain" on developmental genes (Bernstein et al. 2006; Mikkelsen et al. 2007). In both ESCs and human T cells, the relative levels of H3K4me3/H3K27me3 correlate with the state of transcription (Barski et al. 2007; Mikkelsen et al. 2007),

H1 has been shown to interact with a number of proteins responsible for specific histone modifications. For example, H1 interacts with components of the PRC2 complex (Martin et al. 2006), which catalyzes di- or tri- methylation of H3K27 (Cao et al. 2002; Shen et al. 2008; Margueron and Reinberg 2011). This interaction also stimulates PRC2 enzymatic activity (Martin et al. 2006). H1 has also been shown to interact with HDAC SirT1 to induce histone deacetylation and transcription repression. H1 interacts, both *in vitro* and *in vivo*, with HP1 (Nielsen et al. 2001; Daujat et al. 2005; Hale et al. 2006), the key factor involved in establishment and maintenance of heterochromatin structure (Eissenberg and Elgin 2000; Campos and Reinberg 2009). HP1 interacts with H3K9 methylation and H3K9me3 catalyzing enzyme Suv39h1 (Bannister et al. 2001; Lachner et al. 2001). Furthermore, recent reports demonstrated that H1 interacts with *Drosophila* Su(var)3-9 and that *Drosophila* H1 regulates the distribution and level of H3K9

methylation (Lu et al. 2009b; Lu et al. 2013). These results suggest a correlation of H1 with H3K9me3 and constitutive heterochromatin.

Core histone modifications mark regulatory sequences and reflect chromatin structures. Comparing the genome-wide distribution of key histone modifications with that of H1 variants would facilitate elucidating mechanisms of H1 regulated gene expression and chromatin conformation.

1.5 H1 Linker Histone and DNA Methylation

Previous studies suggest that H1 levels have an impact on DNA methylation status either globally or at specific loci in different organisms. Silencing H1 in fungi *Ascobolus immersus* leads to DNA hypermethylation globally (Barra et al. 2000), whereas knockdown of H1 in *Arabidopsis thaliana* leads to stochastic changes of DNA methylation at multiple loci throughout the genome (Wierzbicki and Jerzmanowski 2005). Recent studies in H1 TKO mouse ESCs showed that the regulatory regions of several H1 target genes are hypomethylated, suggesting DNA methylation as one of the mechanisms by which H1 regulates specific gene expression (Fan et al. 2005; Giambra et al. 2008; Maclean et al. 2011; Yang et al. 2013).

DNA methylation plays important roles in repetitive elements silencing, X chromosome inactivation, genomic imprinting and transcription regulation (Li et al. 1993; Beard et al. 1995; Walsh et al. 1998; Robertson and Wolffe 2000; Suzuki and Bird 2008; Bartolomei 2009; Chow and Heard 2009). In mammals, there are three characterized enzymes that are responsible for maintenance and establishment of DNA methylation,

DNMT1, DNMT3a and DNMT3b (Jurkowska et al. 2011). In mouse embryos and mouse embryonic fibroblasts (MEF), depletion of the maintenance methyltransferase DNMT1 leads to reduction of methylation and aberrant expression of imprinted genes, among which several genes display expression alterations in H1 TKO ESCs (Li et al. 1993; Howell et al. 2001; Jackson-Grusby et al. 2001; Fan et al. 2005). Similarly, loss of *de novo* methyltransferases DNMT3a and DNMT3b results in upregulation of H1-regulated Rhox genes in mouse ESCs (Maclean et al. 2011). Interestingly, both DNMT3a and DNMT3b have been shown to bind tightly to nucleosomes and condensed chromatin (Jeong et al. 2009; Kashiwagi et al. 2011; Sharma et al. 2011). Furthermore, DNMT1 and DNMT3b can interact with H1 *in vitro* and *in vivo* in ESCs (Kashiwagi et al. 2011; Yang et al. 2013).

Taken together, these results suggested that loss of H1 disrupts the machinery of DNA methylation at specific loci in mouse ESCs and that loss of recruitment of DNMTs could be one of the mechanisms involved.

1.6 H1 Linker Histone and Embryonic Stem Cells

Embryonic stem cells (ESCs) are pluripotent cells derived from *in vitro* culture of mouse blastocysts (Evans and Kaufman 1981). They have the ability to self-renew and differentiate into cell types from all three germ layers (the endoderm, the mesoderm, and the ectoderm), providing promising resources for regenerative medicine as well as ideal tools for studying animal development and cell differentiation. Mouse ESCs are also broadly utilized in generating mouse models carrying desired genotypes (Mansour et al.

1988) due to the capability of germline transmission (Bradley et al. 1984) (Figure 1.4). These mouse models are proved to be valuable systems for biomedical research and pharmaceutical applications.

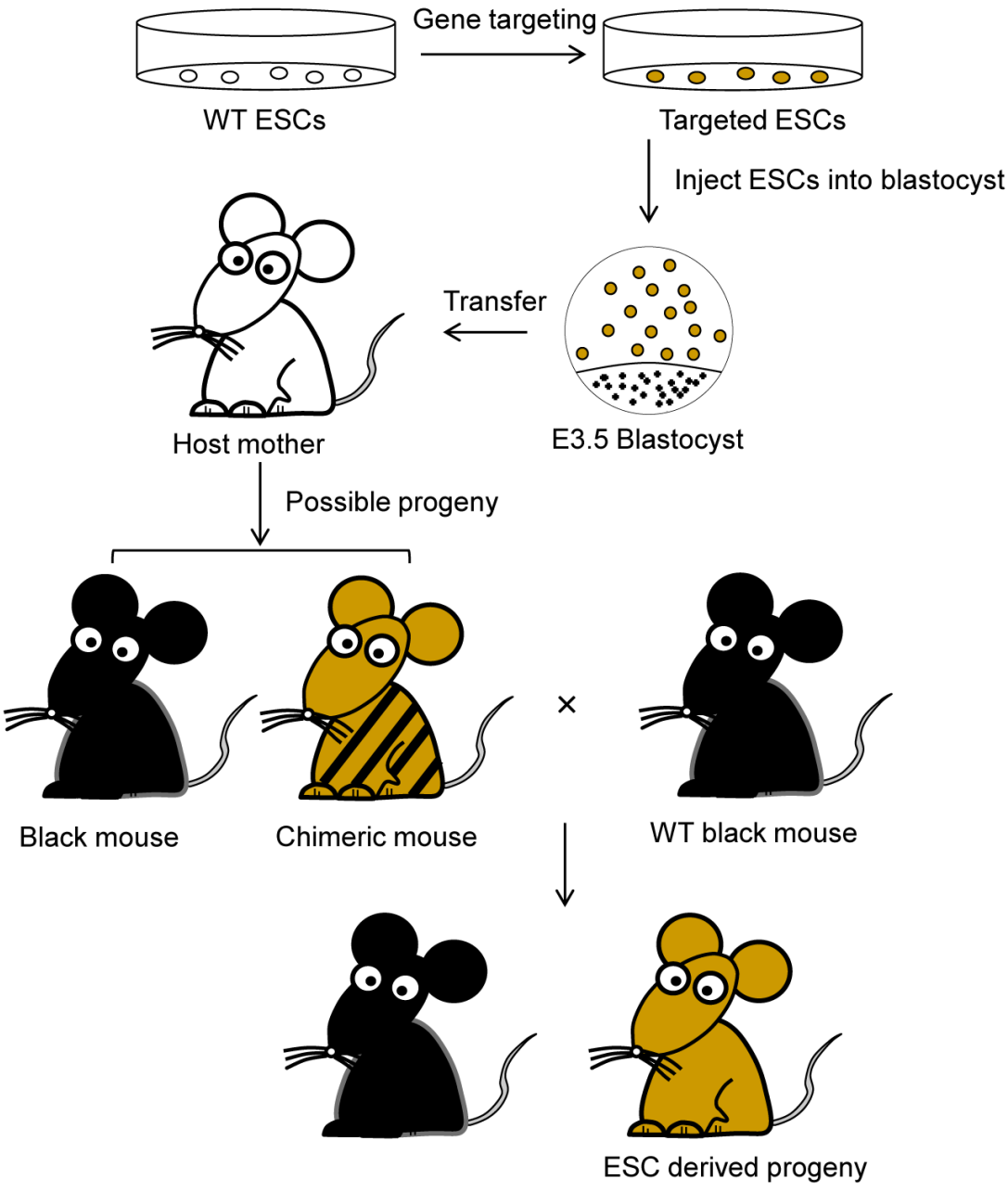

Figure 1.4 The strategy of generating ESC derived mouse.

ESC pluripotency is governed by complex networks of transcription factors and chromatin proteins, such as the ESC core transcriptional circuitry, DNA modifying enzymes, histone modifying enzymes, and chromatin remodelers (Orkin and Hochedlinger 2011; Young 2011). Previous reports suggest that ESCs harbor characteristic transcription and chromatin state, with globally elevated transcription levels and more "open" chromatin compared with differentiated cells (Meshorer et al. 2006; Efroni et al. 2008; Ahmed et al. 2010). The epigenome of ESCs has been extensively characterized in recent years (Bernstein et al. 2006; Guenther et al. 2007; Mikkelsen et al. 2007; Meissner et al. 2008; Ernst et al. 2011), revealing the essential role of epigenetic marks in regulating pluripotency and cell fate determination. However, the genome-wide localization of H1 linker histone, one of the major proteins participating in chromatin compaction and formation of higher order chromatin structure, remains unexplored in the ESC genome.

Embryos lacking three somatic H1 variants die at mid-gestation (Fan et al. 2003), while H1 TKO ESCs are viable and appear similar to WT ESCs in growth and morphology (Fan et al. 2005). Compared with WT ESCs, H1 TKO ESCs display chromatin decondensation, changes of two histone modifications in bulk chromatin, as well as alterations in the expression levels and DNA methylation at specific gene loci (Fan et al. 2005; Giambra et al. 2008; Maclean et al. 2011; Yang et al. 2013). Interestingly, overexpression of a mutant $H1^0$, with two C-terminal domains and higher chromatin binding affinity, in ESC leads to differentiation arrest, suggesting that a proper chromatin binding affinity of $H1^0$ is necessary for cell differentiation (Meshorer et al. 2006). Recent results have shown that triple H1 deletion impairs embryoid body (EB)

differentiation and blocks neural differentiation of ESCs, further emphasizing the requirement of H1 and chromatin compaction for ESC pluripotency (Zhang et al. 2012a).

In summary, H1 linker histone plays an important role in transcription regulation and is required for ESC pluripotency. New tools for studying H1 proteins would provide insights for understanding the important epigenetic regulation by H1 and chromatin compaction in ESCs.

1.7 Objectives

In this study, I aim to investigate the genome-wide localization of H1 variants, correlate the binding patterns of H1 variants with the well studied ESC epigenome, and elucidate the mechanisms of H1 regulated transcription in ESCs. To this end, I have generated multiple cell lines and performed ChIP-seq to map three H1 variants in ESCs. The results integrated H1 to the ESC epigenome and identified novel functions of this important chromatin protein. The cells and mice generated are valuable resources for future investigations on H1 linker histone and higher order chromatin structure.

CHAPTER 2

GENERATION OF TOOLS FOR GENOME-WIDE PROFILING OF H1 VARIANTS IN MOUSE ESCS

The results in Chapter 2 have been published in the below article:

Cao K., Lailler N., Zhang Y., Kumar A., Uppal K., Liu Z., Lee E. K., Wu H., Medrzycki M., Pan C., Ho P. Y., Cooper G. P. Jr., Dong X., Bock C., Bouhassira E. E., Fan Y. (2013) High-Resolution Mapping of H1 Linker Histone Variants in Embryonic Stem Cells. PLoS Genet. 9(4): e1003417

2.1 Abstract

H1 linker histones bind to and stabilize the nucleosome core particle, facilitating higher order chromatin folding. Multiple H1 variants exist in mammals, while their roles in gene regulation and chromatin organization remain elusive. Elucidating the *in vivo* functions of H1 variants would provide critical clues to the interplay among H1 and other epigenetic regulatory mechanisms, however, it is hindered because antibodies of high quality to differentiate individual variants are not available. Here, we have established a knock-in system, in which endogenous H1 variants are replaced with N-terminally tagged variants in ESCs and mice. We have shown that the tagged H1 proteins have the same biochemical properties as endogenous H1s. The tagged H1d could rescue the lethal phenotype of triple H1 depletion in a mouse model. These results indicate that the tagged H1 proteins are functionally interchangeable to their endogenous counterparts *in vivo*. In addition, ESCs containing overexpressed tagged $H1^0$ are generated to facilitate the characterization of this replacement H1 variant in mESCs. The cell lines and mice generated provide useful resources for studying H1 variants *in vivo*.

2.2 Introduction

To date, 11 different H1 variants have been characterized in mammals (Happel and Doenecke 2009). Multiple variants add to the complexity and possibility in regulation of chromatin compaction. Deletion of three major somatic H1 variants (H1c, H1d and H1e) together leads to a 50% reduction of the total H1 level and embryonic lethality at midgestation, demonstrating that H1 level is critical for mammalian development (Fan et al. 2003). H1 variants are conserved from mouse to human, and differ in their biochemical properties and expression patterns during development and malignant transformation (Lennox and Cohen 1983; Wang et al. 1997; Warneboldt et al. 2008; Medrzycki et al. 2012b), suggesting that individual variants are functionally distinct. Mapping of the precise genomic localizations of different H1 variants *in vivo* is likely to provide significant insights, but has been challenging due to the lack of high quality antibodies that could accurately distinguish different H1 variants.

Embryonic stem cells (ESCs) are pluripotent cells that could be differentiated into any cell types of the body, offering great models for studying development and disease. The epigenome plays a critical role in stem cell fate determination, and genome-wide mapping studies have revealed that ESCs have characteristic epigenetic landscapes that differ from differentiated cells (Lu et al. 2009a; Meissner 2010). H1 governs the extent of chromatin compaction, regulates the levels of transcription and DNA methylation at specific genes, and impacts the differentiation potentials, of ESCs (Fan et al. 2005; Zhang et al. 2012a). Despite significant efforts to characterize the chromatin features of human and mouse ESCs, both by individual labs (Bernstein et al. 2006; Creyghton et al. 2008;

Lu et al. 2009a; Creyghton et al. 2010) and by large consortia (ENCODE (Birney et al. 2007), Roadmap Epigenomics (Bernstein et al. 2010)), the landscapes of linker histone H1 variants have not been described on a genome-wide scale largely due to the lack of tools to analyze these variants.

Here, we have generated FLAG-tagged H1d knock-in ESCs, Myc-tagged H1c knock-in ESCs, as well as FLAG-tagged $H1^0$ overexpressing ESCs, designated as respective $H1d^{FLAG}$, $H1c^{Myc}$, and $fH1^0$ mESCs. H1d and H1c are among the most abundant linker histones in mouse ESCs, accounting respectively for 32.6% and 16.4% of total H1, whereas the differentiation associated H1, $H1^0$, accounts for 2% of H1 in undifferentiated ESCs (Fan et al. 2005; Zhang et al. 2012b). These three variants differ significantly in terms of their residence time on chromatin and their ability to promote chromatin condensation *in vitro* (Th'ng et al. 2005; Clausell et al. 2009). They also display different expression patterns during mammalian development and in exponentially growing cells *vs.* quiescent cells (Lennox and Cohen 1983; Meergans et al. 1997; Wang et al. 1997). We demonstrate that tagged H1 variants maintain the biochemical properties of the endogenous H1s *in vivo* and that FLAG-H1d can substitute for H1d during mouse development.

2.3 Materials and Methods

2.3.1 Cell Culture

ESCs were expanded on mitotically inactivated mouse embryonic fibroblasts feeder layers and cultured feeder-free on tissue culture-treated dishes (Corning) pre-adsorbed with 0.1% gelatin (Sigma-Aldrich). ESC culture media consisted of Dulbecco's modified Eagle's medium (DMEM) (Life Technologies) supplemented with fetal bovine serum (FBS) (Gemini), penicillin and streptomycin (Life Technologies), MEM non-essential amino acids (Life Technologies), β-mercaptoethanol (Life Technologies), and leukemia inhibitory factor (LIF; ESGRO, Chemicon).

2.3.2 Generation of $H1d^{FLAG}$ ESCs and $H1d^{FLAG/FLAG}$ mice

The $H1d^{FLAG}$ knock-in targeting vector containing H1d 5' and 3' homology regions flanking the N-terminal FLAG-tagged H1d and the SV40-Blasticidin resistant gene was transfected into ESCs as described previously (Fan et al. 2001). 200 ESC clones resistant to 20 μg/ml Blasticidin (Life Technologies) and 2 μM gancyclovir (Sigma-Aldrich) were picked, and 5 clones with homologous recombination were identified by Southern blotting using the probe shown in Figure 2.1B. Two *cis*-targeted clones were injected into C57BL/6 recipient blastocysts to produce chimeric mice, which gave germline transmission. $H1c^{+/-}H1d^{+/FLAG}H1e^{+/-}$ mice were intercrossed to generate $H1c^{-/-}H1d^{FLAG/FLAG}H1e^{-/-}$ ($H1d^{FLAG/FLAG}$) mice. All animal work was performed according to procedures approved by the Institutional Animal Care and Use Committee (IACUC) at Georgia Institute of Technology.

2.3.3 Generation of H1c^{Myc} ESCs

The H1c^{Myc} knock-in targeting vector containing H1c 5' and 3' homology regions flanking the N-terminal Myc-tagged H1c and the SV40-Blasticidin resistant gene was transfected into ESCs. Cells were selected at the conditions as selecting H1d^{FLAG} ESC clones. 200 ESC clones were picked and 3 *cis*-targeted clones were identified.

2.3.4 Generation of FLAG-H1^0 overexpressing (fH1^0) ESCs

The H1^0 overexpressing vector was constructed by cloning the H1d regulatory regions flanking N-terminal FLAG-tagged H1^0 into a plasmid containing SV40-Blasticidin resistant gene. Tranfection was performed in WT ESCs and 48 cell clones resistant to 20 μg/ml Blasticidin (Life Technologies) were picked. Two cell lines with the highest levels of H1^0 were selected for further analysis.

2.3.5 Preparation and analysis of histones

Nuclei and chromatin of ESCs and mouse tissues were prepared and analyzed according to protocols described previously (Fan and Skoultchi 2004; Medrzycki et al. 2012a). Histones were extracted from chromatin with 0.2 N sulfuric acid and 50-100 μg of total histone preparations were injected into a C18 reverse phase column (Vydac) on an ÄKTA UPC10 system (GE Healthcare). The effluent was monitored at 214 nm (A_{214}), and the peak areas were recorded and analyzed with ÄKTA UNICORN 5.11 software. The A_{214} values of the H1 and H2B peaks were adjusted by the number of peptide bonds in each H1 variant and H2B. The H1/nucleosome ratio was determined by dividing the A_{214} of all H1 peaks by half of the A_{214} of the H2B peak. Fractions corresponding to

different H1 variants from HPLC analysis were collected, lyophilized and analyzed with silver staining, Coomassie staining and Western blotting.

2.3.6 Antibodies

The following antibodies were used in this study: anti-FLAG (Sigma-Aldrich F3165), anti-Myc-tag (Cell Signaling #2272), anti-$H1^0$ (Santa Cruz 56695), and anti-H1 (Milipore 05-457).

2.4 Results

2.4.1 Generation of tagged H1d knock-in mESCs

Efforts to generate high resolution genome-wide maps of H1 variants were hampered by the lack of H1 variant specific antibodies of sufficient quality for ChIP-seq. Here, we established knock-in mouse ESC lines in which H1d or H1c variant was N-terminally tagged with an epitope (FLAG or Myc) for which highly specific antibodies exist. An $H1d^{FLAG}$ cell line was created by inserting a FLAG tag sequence at the start of endogenous H1d coding sequence through homologous recombination (Figure 2.1A). H1c/H1e double knockout mice develop normally, yet H1c/H1d/H1e triple knockout (H1 TKO) mice are embryonic lethal (Fan et al. 2003). Thus, ESCs with $H1d^{FLAG}$ allele in $H1c^{+/-}H1e^{+/-}$ background could be used to produce $H1c^{-/-}H1d^{FLAG/FLAG}H1e^{-/-}$ mice to determine whether FLAG-tagged H1d (FLAG-H1d) functions equivalently to endogenous H1d by assessing if the tagged H1d can rescue the embryonic lethality of H1 TKO mutants. Toward this end, we generated both $H1c^{+/-}H1d^{+/FLAG}H1e^{+/-}$ ("$H1d^{FLAG}$") and $H1c^{+/-}H1d^{FLAG/-}H1e^{+/-}$ ("H1d-*trans*") ESC lines by transfection of the FLAG-H1d targeting vector (Figure 2.1A) into the *cis* triply targeted $H1c^{+/-}H1d^{+/-}H1e^{+/-}$ ESCs established previously (Fan et al. 2003). ESC clones with either *cis* or *trans* configuration of the $H1d^{FLAG}$ allele with the H1c and H1e KO allele were identified and verified by Southern blotting (Figure 2.1B). As expected, FLAG-H1d was located in the nuclei of the $H1d^{FLAG}$ cells (data not shown). Analysis of histone extracts of chromatin prepared from *cis*-targeted $H1d^{FLAG}$ cells by HPLC and immunoblotting indicated that FLAG-H1d was associated with chromatin and co-eluted in the same fraction as the

endogenous H1d, suggesting that FLAG-H1d has the same hydrophobicity as the endogenous H1d (Figure 2.2A and 2.2B). The ratio of somatic H1 variants, H1 a-e, to nucleosome (H1/nuc) of $H1d^{FLAG}$ cells was nearly identical to that of $H1c^{+/-}H1d^{+/+}H1e^{+/-}$ (ce^{het}) cells, indicating a similar expression level of FLAG-H1d as the endogenous H1d (Figure 2.2C). As expected, the protein level of differentiation associated $H1^0$ variant is minimal in undifferentiated ESCs.

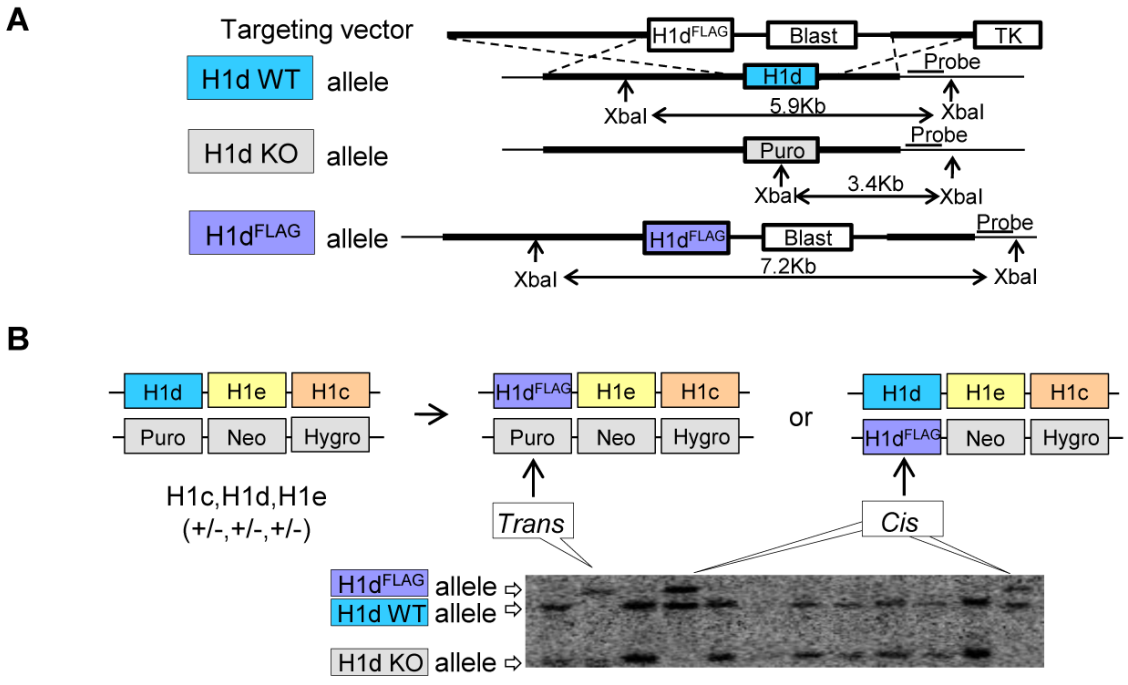

Figure 2.1 Generation of H1d^{FLAG} knock-in ESCs.

(A) Schematic representation of the H1d^{FLAG} targeting construct and the knock-in strategy for insertion of the FLAG tag at N-terminus of the endogenous H1d gene.

(B) Identification of ESC clones containing the modified FLAG-H1d allele. DNA isolated from Blasticidin resistant ESC clones were analyzed by Southern blotting. *Cis vs. trans* configurations of the homologous recombination events are schematically illustrated in the diagram above the Southern blotting image.

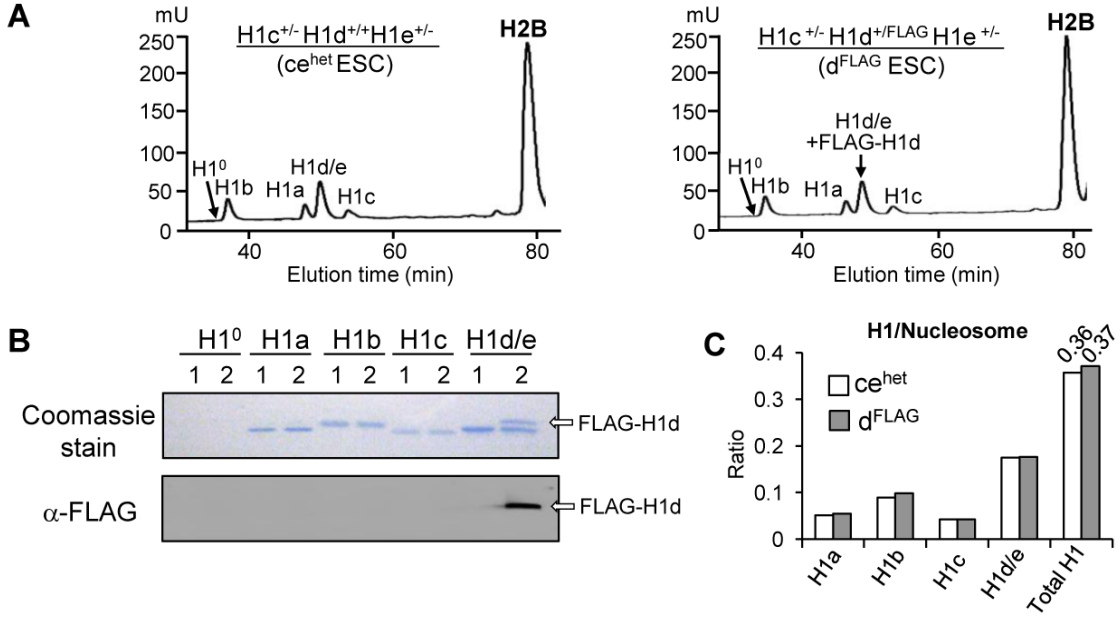

Figure 2.2 FLAG-H1d has similar biochemical properties as endogenous H1d.

(A) Reverse phase HPLC profiles of histone extracts from ce^{het} (left panel) and *cis*-targeted H1d^{FLAG} ESCs (right panel). mU, milliunits of absorbency at 214 nm.

(B) Coomassie staining and Western blotting analysis of individual H1 fractions eluted from HPLC of histone extracts of ce^{het} (1) and H1d^{FLAG} (2) ESCs.

(C) Calculated ratio of each H1 variant (and total H1) to nucleosome of ce^{het} and H1d^{FLAG} ESCs.

2.4.2 Generation of tagged H1d knock-in mice

To generate chimeric mice which gave germline transmission of the $H1d^{FLAG}$ allele, I collaborated with Yunzhe Zhang, who injected *cis*-targeted $H1d^{FLAG}$ ESCs into mouse blastocysts, as well as Dr. Zheng Liu, who transferred the blastocysts into pseudopregnant mice. $H1c^{+/-}H1d^{+/FLAG}H1e^{+/-}$ mice were intercrossed to generate $H1c^{-/-}H1d^{FLAG/FLAG}H1e^{-/-}$ homozygous mice (designated as $H1d^{FLAG/FLAG}$ mice) (Figure 2.3A). These homozygotes were viable, fertile and developed normally as H1c/H1e double null (ce^{KO}) mice, demonstrating that FLAG-H1d can substitute for the endogenous H1d to fully rescue the lethal phenotype of H1 TKO mutants. HPLC, mass spectrometry and immunoblotting demonstrated that $H1d^{FLAG/FLAG}$ mice had full replacement of H1d by FLAG-H1d (Figure 2.3B and Figure 2.4) and that the H1/nuc ratio of spleen chromatin from $H1d^{FLAG/FLAG}$ mice was 0.7, comparable to that of ce^{KO} mice (Figure 2.5). Taken together, these results demonstrate that FLAG-H1d maintains the expression level and properties of the endogenous H1d *in vivo*.

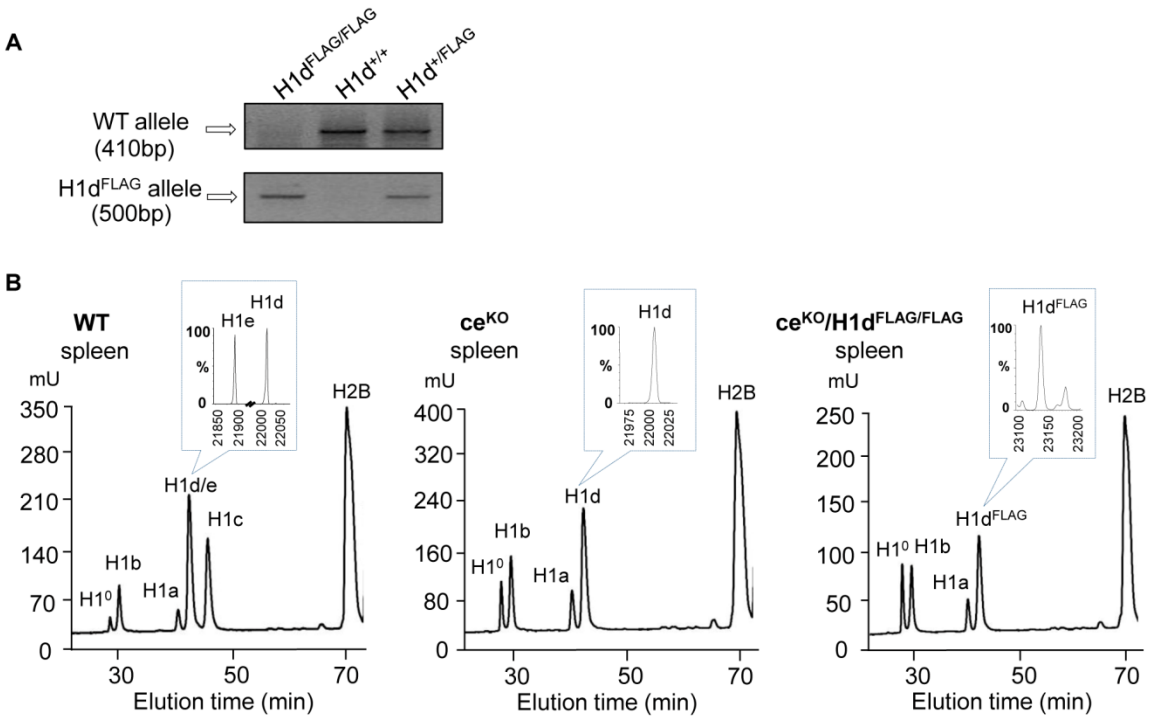

Figure 2.3 Generation and HPLC analysis of H1d$^{FLAG/FLAG}$ mice.

(A) Genotyping analysis of H1d$^{FLAG/FLAG}$ mice. The positions of WT and H1d^{FLAG} PCR bands are indicated by arrows.

(B) Reverse phase HPLC and mass spectrometry analysis of extracted histones from spleens of 1-year-old wildtype (WT, left), H1c$^{-/-}$ H1d$^{+/+}$ H1e$^{-/-}$ (ce^{KO}, middle), and H1c$^{-/-}$ H1d$^{FLAG/FLAG}$ H1e$^{-/-}$ mice (ce^{KO}/H1d$^{FLAG/FLAG}$, right). The insets are profiles generated by ESI-TOF mass spectrometry analysis of H1d/e fraction eluted from HPLC.

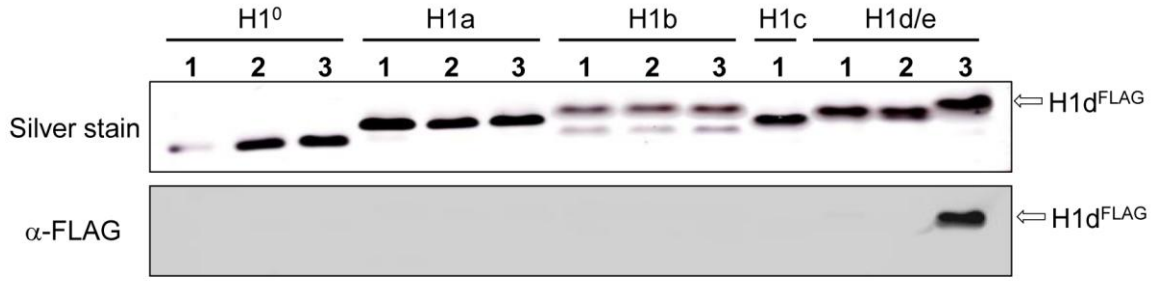

Figure 2.4 Silver staining and immunoblotting assays of H1 variants from mouse spleen.

Individual H1 variants were eluted from HPLC analysis shown in Figure 2.3B. 1: WT, 2: ce^{KO}, 3: ce^{KO}/H1d$^{FLAG/FLAG}$.

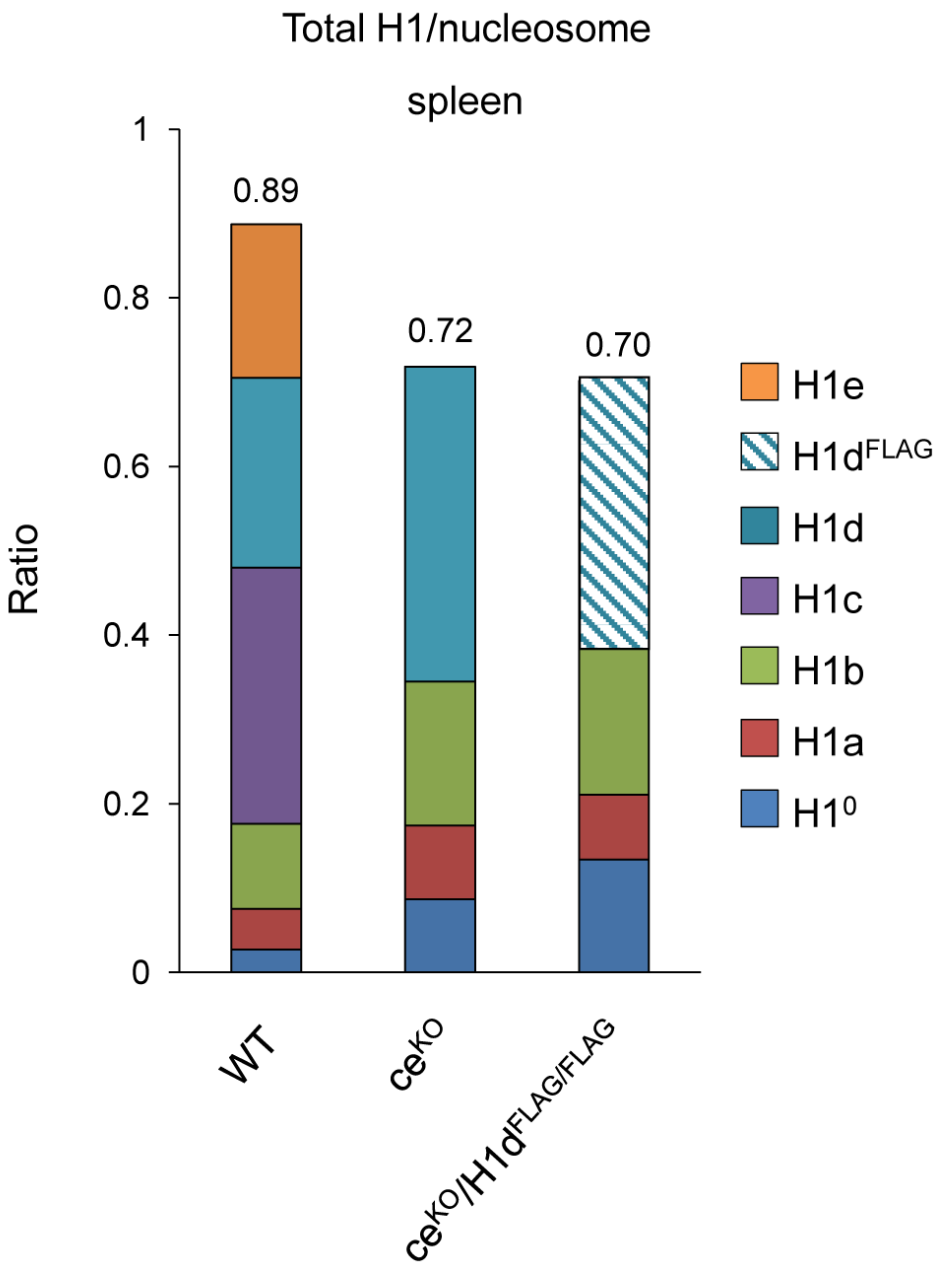

Figure 2.5 H1/nucleosome ratio of histone extracts from mouse spleen.
Values were calculated from HPLC analysis as shown in Figure 2.3B.

2.4.3 Generation of tagged H1c knock-in mESCs

Using a similar knock-in strategy, we generated $H1c^{+/Myc}H1d^{+/-}H1e^{+/-}$ ESCs ($H1c^{Myc}$) by transfection of the $H1c^{Myc}$ targeting construct into the *cis* triply targeted $H1c^{+/-}H1d^{+/-}H1e^{+/-}$ ESCs and selected ESC clones that underwent homologous recombination at H1c locus (Figure 2.6A and 2.6B). Similar to FLAG-H1d, the N-terminally Myc tagged H1c (Myc-H1c) colocalized with Hoechst stained nuclear regions in $H1c^{Myc}$ cells (data not shown), and Myc-H1c was eluted in the same fraction as the endogenous H1c protein from HPLC analysis (Figure 2.7A and 2.7B). $H1c^{Myc}$ cells had a H1/nuc ratio of 0.38, comparable to the ratio of 0.36 in ce^{het} cells (Figure 2.2C, Figure 2.7A), indicating that like FLAG-H1d, Myc-H1c has the same expression level and biochemical properties as the endogenous H1c.

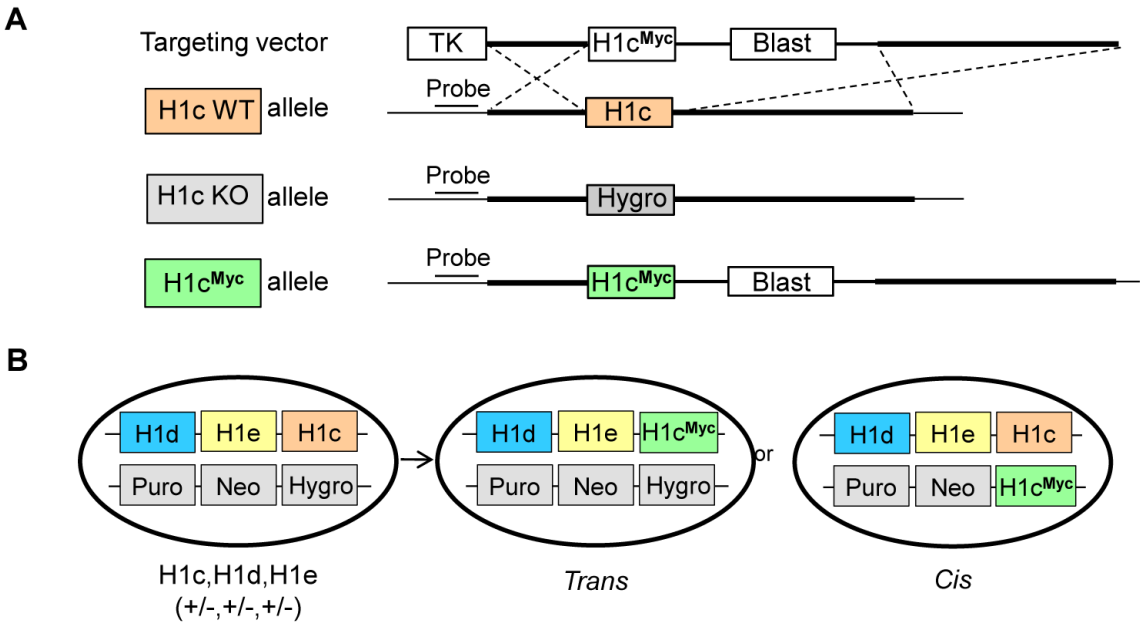

Figure 2.6 Generation of H1c^{Myc} knock-in ESCs
(A) Schematic representation of the H1c^{Myc} targeting vector and homologous recombination which results in insertion of the Myc tag at N-terminus of the coding sequence of the endogenous H1c gene.
(B) Strategy of constructing H1c^{Myc} knock-in ESCs and *cis vs. trans* configurations of the homologous recombination events.

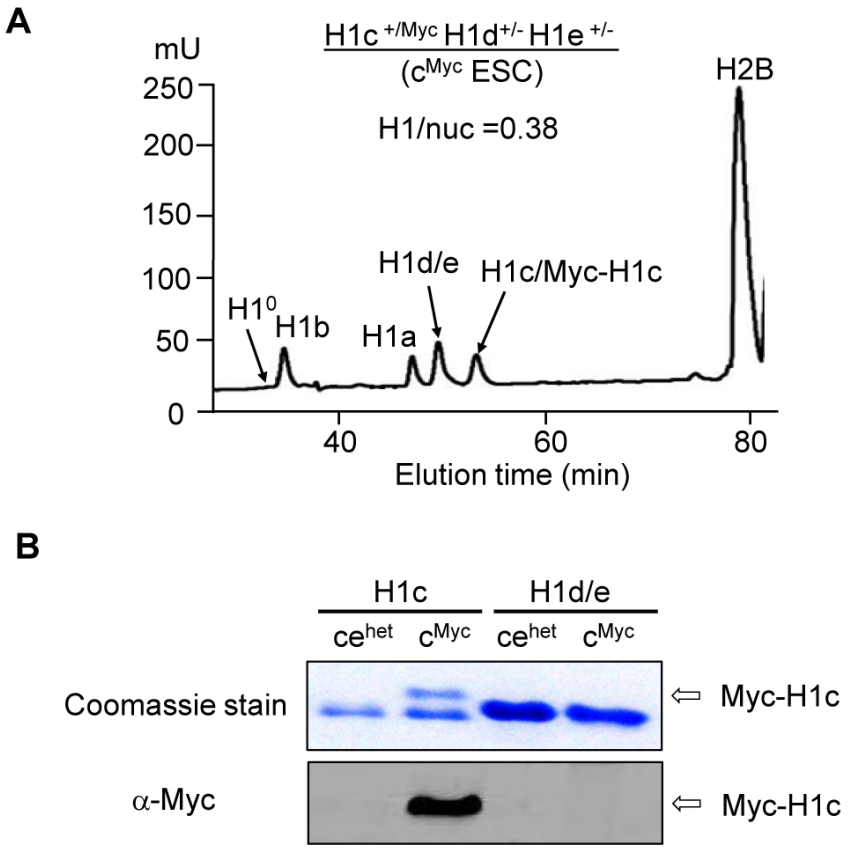

Figure 2.7 Anlaysis of hisone extracts from H1c^{Myc} knock-in ESCs.
(A) Reverse phase HPLC analysis of total histone extracts from H1c^{Myc} cells.
(B) Coomassie stain (top) and immunoblotting (bottom) assay of the H1c and H1d/e peaks eluted from HPLC of histone extracts from ce^{het} cells and H1c^{Myc} cells.

2.4.4 Generation of tagged $H1^0$ overexpressing mESCs

The replacement variant $H1^0$ is more divergent in amino acid sequence compared with the somatic variants, and is highly expressed in terminally differentiated tissues (Zlatanova and Doenecke 1994; Happel and Doenecke 2009), suggesting a possible different chromatin binding pattern. To examine the genome localization of $H1^0$, we generated "$fH1^0$" cells by over-expressing FLAG-$H1^0$ in WT ESCs (Figure 2.8), and selected cell lines that expressed FLAG-$H1^0$ at a similar level to that of $H1^0$ in H1 TKO ESCs (Figure 2.9, Figure 2.10 and (Fan et al. 2005)). As expected, FLAG-$H1^0$ was eluted in the same fraction as endogenous $H1^0$ (Figure 2.9), demonstrating the similar biochemical properties of tagged and endogenous $H1^0$.

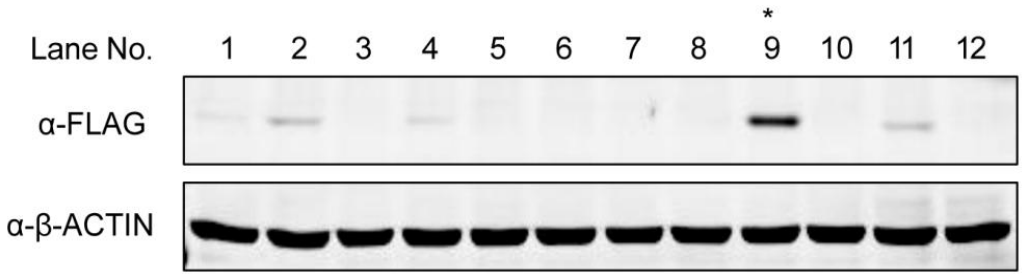

Figure 2.8 Generation of $H1^0$ over-expressing ($fH1^0$) ESCs.

Representative Western blots of $H1^0$ overexpressing cell clones. WT ESCs were transfected with the vector expressing FLAG-$H1^0$, and stable ESC clones were picked and screened using an anti-FLAG antibody. Immunoblotting with anti-β-ACTIN antibody indicates equal loading of whole cell lysates. An $H1^0$ overexpressing clone with significant levels of FLAG-$H1^0$ is indicated with an asterisk.

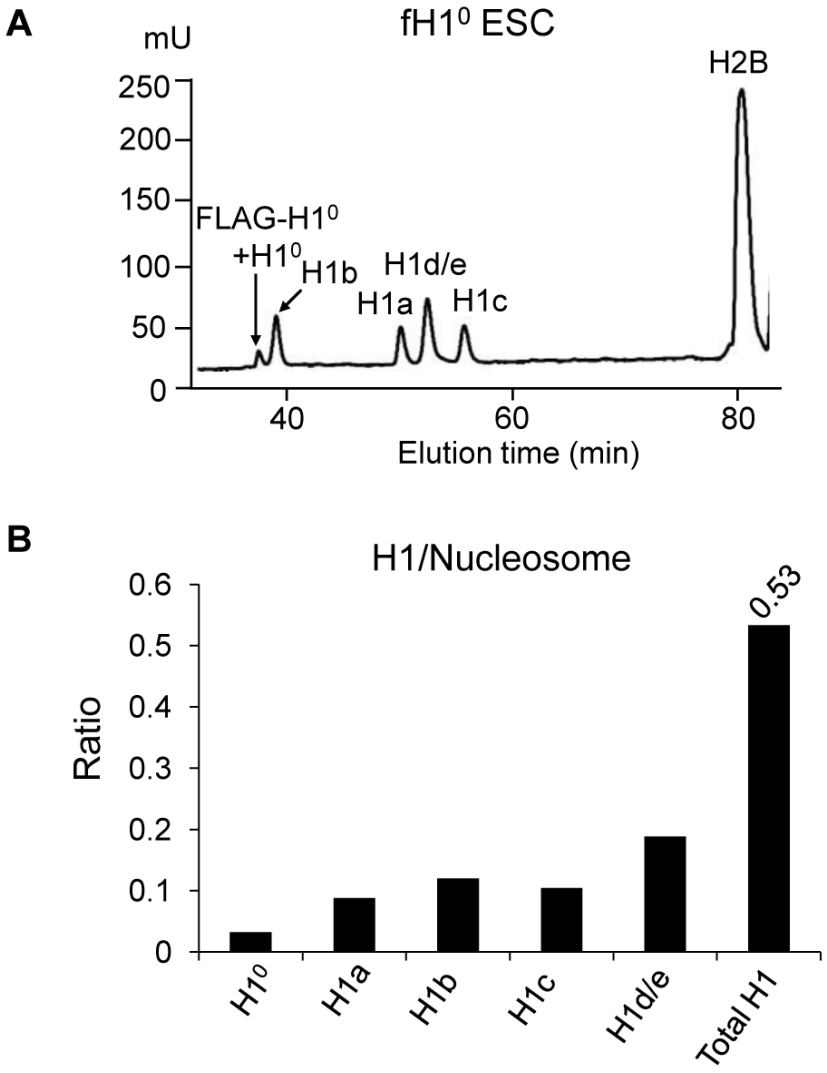

Figure 2.9 HPLC analysis of fH1^0 ESCs.

(A) RP-HPLC Profile of fH1^0 ESCs.

(B) Ratio of individual H1 variant (and total H1) to nucleosome of fH1^0 ESCs calculated from HPLC profile shown in (A).

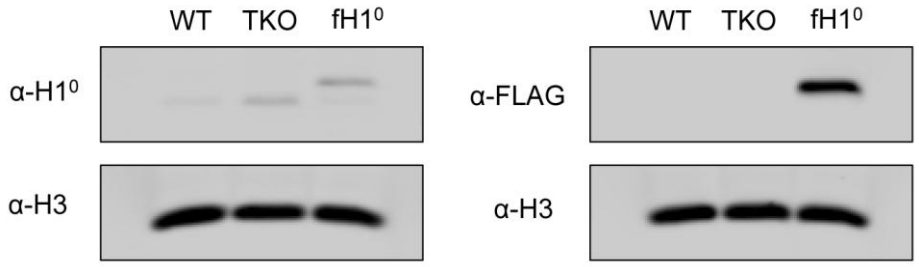

Figure 2.10 Western blots indicating similar levels of $H1^0$ in H1 TKO cells and the FLAG-$H1^0$ in $fH1^0$ ESCs.
H3 blots indicates equal loading of chromatin lysates.

2.5 Discussion

Deletion of multiple H1 variants leads to embryonic lethality and differentiation impairment in respective mice and ESCs (Fan et al. 2003; Zhang et al. 2012a), suggesting the importance of chromatin compaction during mammalian development and ESC lineage commitment. The similarity and divergence among H1 variants adds to another level of regulation in higher order chromatin structure. Elucidating the *in vivo* functions of individual H1 variants would pave the way for understanding the roles of chromatin state in development and disease, however, lacking of powerful tools such as antibodies against specific variants places impediment for such studies. Here, we have achieved to produce H1 knock-in ESCs and mice as useful systems to map the genome localizations of H1 variants *in vivo*.

We established a knock-in system to stringently test the functions of the tagged H1s and to facilitate the generation of high resolution maps of H1 variants in ESCs by ChIP-seq. Our results demonstrate that, when tagged at the N-terminus, the short FLAG and Myc tags, with respective 8 and 13 amino acids, do not alter the biochemical and cellular properties of H1 proteins *in vivo*. The strategy of homologous recombination ensures that the expression of tagged H1 variants is comparable to that of their endogenous counterparts. FLAG-H1d fully rescues the lethal phenotype of H1d deletion on H1c/H1e double knockout genetic background, further demonstrating the functional equivalence of the tagged H1 and the respective endogenous H1 variant *in vivo*. Although Myc-H1c was not tested in mice, it is anticipated to mimic the endogenous H1c based on all the other assays performed.

The expression of $H1^0$ in WT ESCs is minimal, thus it is challenging to map the localization of the replacement variant by antibodies recognizing the endogenous proteins. Here, FLAG tagged $H1^0$ was overexpressed in WT ESCs to the level similar to that in H1 TKO ESCs, providing an opportunity to study the genome binding pattern of physiological levels of $H1^0$ in undifferentiated cells.

Taken together, the data presented here provide a technical demonstration on how highly similar protein variants can be analyzed differentially using *in vivo* validated knock-in mice. The cell lines and mouse strains generated here also provide valuable tools for studying specific functions of H1 variants both *in vitro* and *in vivo*.

CHAPTER 3

HIGH RESOLUTION MAPPING OF H1D, H1C, AND $H1^0$ IN MOUSE ESCS

The results in Chapter 3 have been published in the below article:

Cao K., Lailler N., Zhang Y., Kumar A., Uppal K., Liu Z., Lee E. K., Wu H., Medrzycki M., Pan C., Ho P. Y., Cooper G. P. Jr., Dong X., Bock C., Bouhassira E. E., Fan Y. (2013) High-Resolution Mapping of H1 Linker Histone Variants in Embryonic Stem Cells. PLoS Genet. 9(4): e1003417

3.1 Abstract

H1 variants have distinct biochemical properties and chromatin binding affinity, however, the precise genome localizations of individual variants are still mysterious. Here, we performed ChIP-seq using the tagged H1 knock-in ESCs and $H1^0$ over expressing ESCs generated in Chapter 2. Our results indicate that H1d and H1c are depleted from GC- and gene-rich regions and active promoters, inversely correlated with H3K4me3, but positively correlated with H3K9me3 and associated with characteristic sequence features. Unexpectedly, both H1d and H1c are significantly enriched at major satellites, the tandem repeats located at the pericentric region of mouse chromosomes. While also depleted at active promoters and enriched at major satellites, overexpressed $H1^0$ displays differential binding patterns in specific repetitive sequences compared with H1d and H1c. These results integrate the localization of the understudied H1 variants into the epigenome map of mouse ESCs.

3.2 Introduction

Although none of the H1 variants tested is essential for mouse development (Sirotkin et al. 1995; Lin et al. 2000; Rabini et al. 2000; Fan et al. 2001), they have been shown to regulate specific gene expression in various cell types (Gunjan and Brown 1999; Alami et al. 2003; Sancho et al. 2008; Happel and Doenecke 2009). On the other hand, significant compensation from the remaining H1s has been found in mice and cells lacking one or multiple H1 variants (Sirotkin et al. 1995; Lin et al. 2000; Fan et al. 2001; Fan et al. 2003; Lin et al. 2004; Fan et al. 2005), suggesting the redundancy of multiple H1 variants. Studying the precise genome localizations of individual H1 variants would clarify the similarity and difference of these variants as well as the underlying mechanisms of H1 regulated transcription.

Chromatin binding affinity of H1 variants has been scrutinized by FRAP analysis and *in vitro* biochemical analysis, and differences of these variants have been observed (Th'ng et al. 2005; Izzo et al. 2008; Clausell et al. 2009). Despite very innovative and informative, these approaches do not have enough resolution to identify the exact localization of H1 proteins or could not completely reflect protein properties *in vivo*. Chromatin immunoprecipitation (ChIP) is a powerful technique to study a specific protein and DNA interaction *in vivo*. Recent advances in the next-generation sequencing technique further invigorate ChIP and expand it to the genome-wide scale with high resolution (ChIP-seq) (Shendure and Ji 2008; Park 2009). The H1 knock-in and overexpression systems we generated in Chapter 2 offer great opportunity to perform

high resolution genome-wide mapping of individual H1 variants in ESCs utilizing ChIP-seq.

In this study, we have achieved high resolution mapping of H1d, H1c and $H1^0$ in ESCs by chromatin immunoprecipitation followed by massive parallel sequencing (ChIP-seq). High resolution mapping reveals that H1d and H1c occupancies are highly correlated, both enriched at AT-rich regions, but also possess different binding specificity. Both H1d and H1c largely co-localize with H3K9me3, but show an inverse correlation with GC% or H3K4me3. Importantly, we discover that H1d and H1c are highly enriched at major satellite elements.

3.3 Materials and Methods

3.3.1 Antibodies

The following antibodies were used in this study: anti-FLAG (Sigma-Aldrich F3165), anti-Myc-tag (Cell Signaling #2272), anti-H3K4me3 (Millipore 07-473), anti-H3K9me3 (Abcam 8898), anti-H3K27me3 (Millipore 07-449), and IgG (Millipore 12-370).

3.3.2 Chromatin immunoprecipitation (ChIP)

ChIP assays were performed as described previously (Fan et al. 2005) with the following modifications: 20 μl of Dynabeads Protein G (Life Technologies) were incubated with 2 μg of antibody for 4 hours, followed by incubation with 40 μg of sonicated soluble chromatin overnight at 4℃. Dynabeads were washed, immunoprecipitates were eluted, and DNA-protein complexes were incubated overnight at 65℃ to reverse crosslinks. DNA was purified with a DNA Isolation column (Qiagen). Input control DNA was prepared from reverse-crosslinked soluble chromatin prior to immunoprecipitation. Quantitatitve PCR on ChIP samples for major satellites, minor satellites, LINE L1, IAP LTR and Hprt was performed with primers published previously (Martens et al. 2005; Boyer et al. 2006). Primer sequences are listed in Table A.1.

3.3.3 Generation of ChIP-seq libraries

The libraries for massive parallel sequencing were prepared with the ChIP-seq Sample Preparation Kit (Illumina) according to the manufacturer's instructions. Briefly,

10 ng of immunoprecipiated DNA or input DNA were end repaired, 3' adenylated and ligated with adapter oligos supplied by the manufacturer. DNA fragments within the range of 120~500 bp were purified following gel electrophoresis and amplified with primers provided by the manufacturer. Library DNA was subsequently purified with a Qiagen DNA Isolation column, quantified and submitted for sequencing.

3.3.4 Sequence reads processing and alignment

Sequencing was performed with Illumina Genome Analyzer II and Illumina HiSeq 2000 systems, and raw sequence reads containing more than 30% of 'N' were removed and adaptor sequences were trimmed. Clean sequences were aligned against mouse genome, mm9 (UCSC website), and 2,669 categories of mammalian repeats from RepBase version 14.07 (Jurka 2000; Jurka et al. 2005) using Bowtie aligner software (http://bowtie-bio.sourceforge.net/index.shtml). The first 40 bp (for alignment to mm9) or the first 35 bp (for alignment to RepBase) of the reads were used as seed sequences with up to two mismatches allowed for the alignment, and aligned number of reads were scored. Reads with multiple alignment positions were mapped randomly to one of the possible position. Reads for each ChIP-seq or input-seq library aligned to mm9 were normalized to 10 million reads, and IP-IN signals were calculated in each 100 bp sliding window by subtraction of normalized read counts per 10 million mappable reads of ChIP-seq library by that of its corresponding input-seq library using GenPlay software (http://genplay.einstein.yu.edu/wiki/index.php/Documentation) (Lajugie and Bouhassira 2011)). Percentage of reads for each repeat mapped to RepBase was calculated by dividing reads mapped to the respective repeat by the total reads in the library, and the

fold enrichment for the respective repeat was subsequently calculated as the ratio of the percent of reads of ChIP-seq library to that of the input-seq library. Read length and read counts of each library are listed in Table 3.1. Representative ChIP-seq libraries with the most sequencing reads mapped to mm9 were utilized for genome browser visualization and metagene analysis, and all replicate ChIP-seq libraries were included in repetitive sequence analysis.

3.3.5 Genome-wide correlation analysis

The sum of signals (IP-IN) for each 1000 bp window (normalized to 10 million reads) was used to calculate the correlation coefficients of H1 variants with GC% and different histone markers. Genome-wide and chromosome-wide correlation coefficients were calculated, and the scatter-plots were generated using Matlab.

3.3.6 Overrepresentation and distribution pattern analysis

Significantly enriched regions were identified using SICER v1.1 (Zang et al. 2009) at the following parameter settings: window size = 200, gap size = 600, E-value = 1000, an effective genome size of 80% of the entire mouse genome, and q-value (FDR) = 0.001. In order to optimize the gap size for H1 variants, the gap size was varied from 0 to 3 times the window size (0, 200, 400, 600) and the best value was chosen according to the criteria as previously described (Zang et al. 2009). Distribution of peak regions relative to gene regions was analyzed by CEAS (Shin et al. 2009). Top 10% of enriched regions for each ChIP-seq library were selected to identify the overrepresented features using EpiGRAPH (http://epigraph.mpi-inf.mpg.de/WebGRAPH/) (Bock et al. 2009). 2214

H1d/H1c common peaks, 1939 H1d unique peaks, 433 H1c unique peaks, 1891 H3K9me3 peaks, 4778 H3K27me3 peaks, and 3446 H3K4me3 peaks were analyzed by EpiGRAPH.

3.4 Results

3.4.1 H1d and H1c are under-represented at GC-, gene- rich regions and depleted at active promoters

To achieve high resolution mapping of H1d and H1c variants in mouse ESC genome, we performed ChIP-seq in *cis*-targeted $H1d^{FLAG}$ and $H1c^{Myc}$ ESCs using anti-FLAG and anti-Myc antibodies, respectively. Input-seq and ChIP-seq libraries were sequenced and aligned to genome through collaboration with Nathalie Lailler and Dr. Eric Bouhassira at Albert Einstein College of Medicine. In each ChIP-seq library, approximately 80-90% of reads were mappable to the mouse genome (mm9) using the Bowtie aligner (Langmead et al. 2009) (Table 3.1). While sonicated chromatin input control libraries on average had 65% *vs.* 22% of reads mapped to unique positions and multiple positions respectively, the H1c ChIP-seq libraries had 44% *vs.* 45% mapped to unique *vs.* multiple positions, suggesting that a higher proportion of H1c resides on repetitive sequences. Similarly, an overrepresentation of multi-match sequence reads (39% of mapped reads) occurred in H1d ChIP-seq libraries. A survey of sequencing signal intensities indicated that H1d and H1c were generally depleted from gene rich regions with the deepest dips around transcription start sites of active genes (examples shown in Figure 3.1, Figure 3.2, Figure 3.3, and Figure 3.4). ChIP-seq with the anti-FLAG antibody in control ESCs not containing FLAG-H1d generated minimal random background signals (data not shown), and examination of H1c (anti-Myc) signals showed no enrichment at c-Myc target genes, such as *Oct4*, *Nanog* and *Sox2* (Kidder et al. 2008) (Figure 3.4), indicating no cross-reactivity for these antibodies. To compare H1

occupancy with other histone marks, we performed ChIP-seq of an active histone mark, H3K4me3, and two repressive histone marks, H3K9me3 and H3K27me3, in murine ESCs. Visual examination of the track files revealed that H1 dips often coincided with H3K9me3 dips or H3K4me3 peaks and that H1 displayed higher signals at gene poor regions with high AT% (low GC%) (Figure 3.1 and Figure 3.2). H3K27me3, enriched at *Hox* gene clusters (Figure 3.5) as expected, did not show obvious pattern correlation with H1 (Figure 3.1, Figure 3.2, Figure 3.3 and Figure 3.4). These observations suggest possible correlations of H1d and H1c with H3K9me3, H3K4me3, gene distribution and GC content *in vivo*.

Table 3.1 List of read length, counts, and total mappable reads (to mm9) of the libraries.

Library	Read length (bp)	Total reads	Mappable reads	% mappable
Input-a	85	33,441,224	29,809,507	89.14
H3K4me3-1	85	22,142,352	19,972,402	90.2
H3K4me3-2	104	29,439,221	24,864,366	84.46
H3K27me3-1	85	36,662,043	31,760,328	86.63
H3K27me3-2	104	30,530,476	27,333,935	89.53
Input-b	104	58,956,901	55,637,627	94.37
H1d-1	104	128,728,185	113,872,952	88.46
H1d-2	104	88,243,326	72,421,297	82.07
H1d-3	100	8,982,438	8,257,555	91.93
H1d-4	100	6,338,749	5,819,605	91.81
H1c-1	104	78,246,028	70,695,286	90.35
H1c-2	100	7,159,846	6,463,193	90.27
H3K9me3-1	104	83,327,029	78,019,097	93.63
H3K9me3-2	100	12,369,001	11,633,045	94.05
H3K9me3-3	100	44,923,673	36,307,313	80.82
H1d-*Trans*-1	100	6,889,836	6,279,397	91.14
H1d-*Trans*-2	100	64,330,869	50,757,056	78.9
$H1^0$	100	42,232,533	37,215,308	88.12

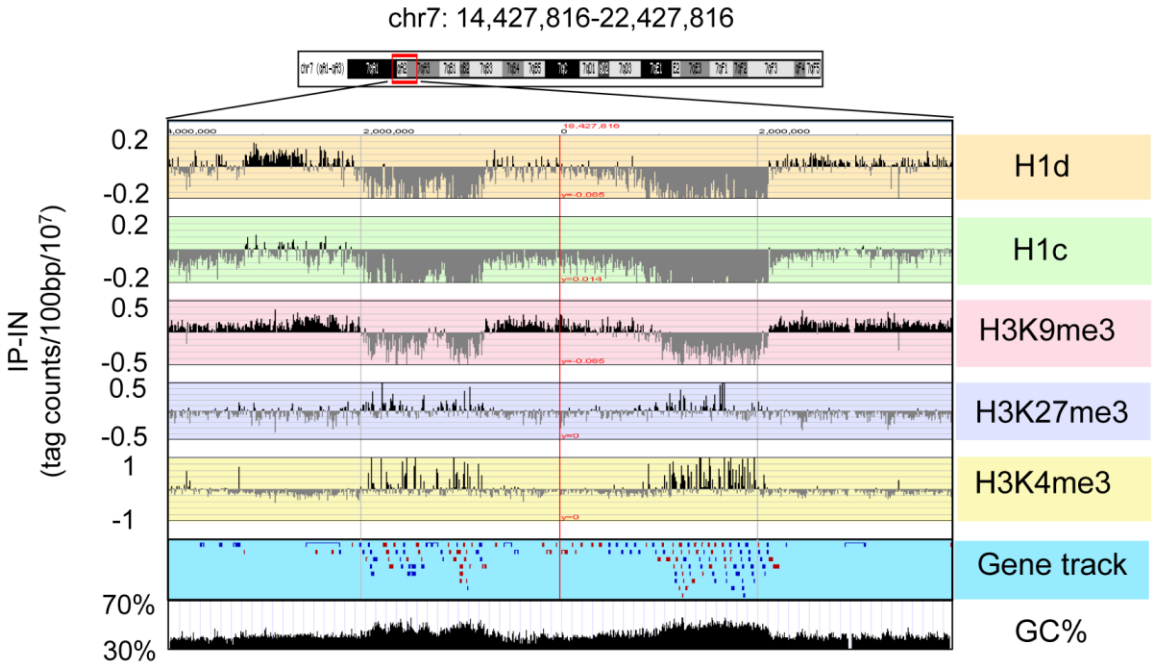

Figure 3.1 Distributions of H1 variants and histone marks at an 8 Mb region.
The GC density track was obtained from UCSC genome browser. Genes are color coded according to their transcription directions (Red: sense strand; Blue: anti-sense strand).

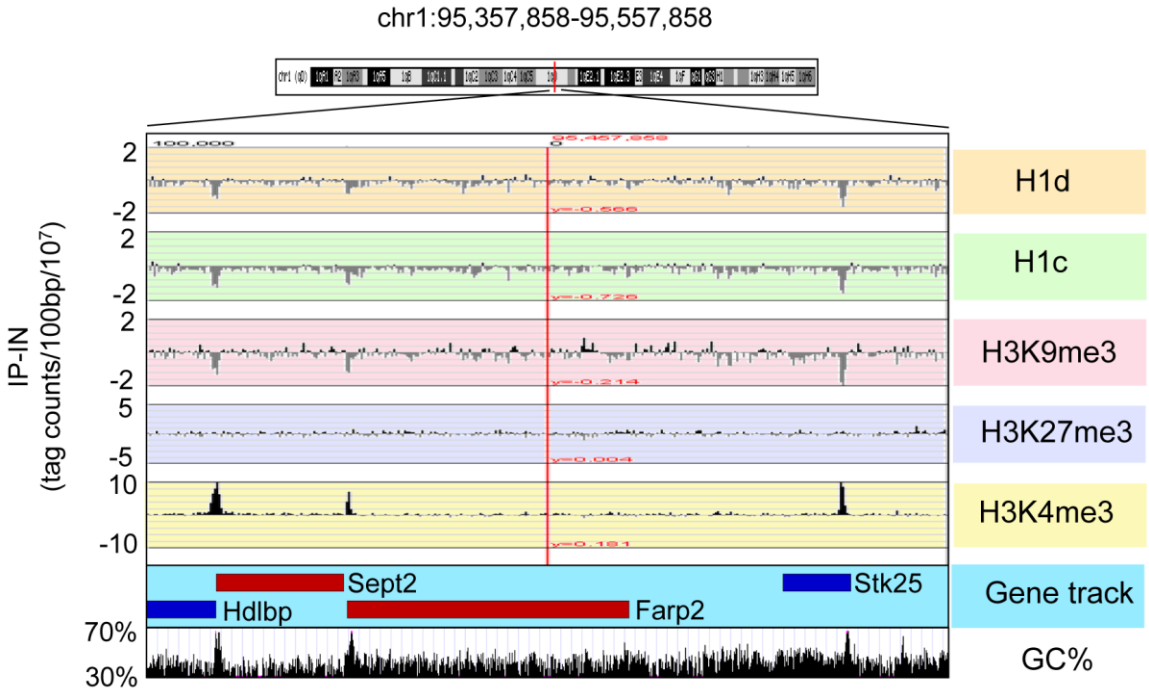

Figure 3.2 Distributions of H1 variants and histone marks at a 200 Kb region.

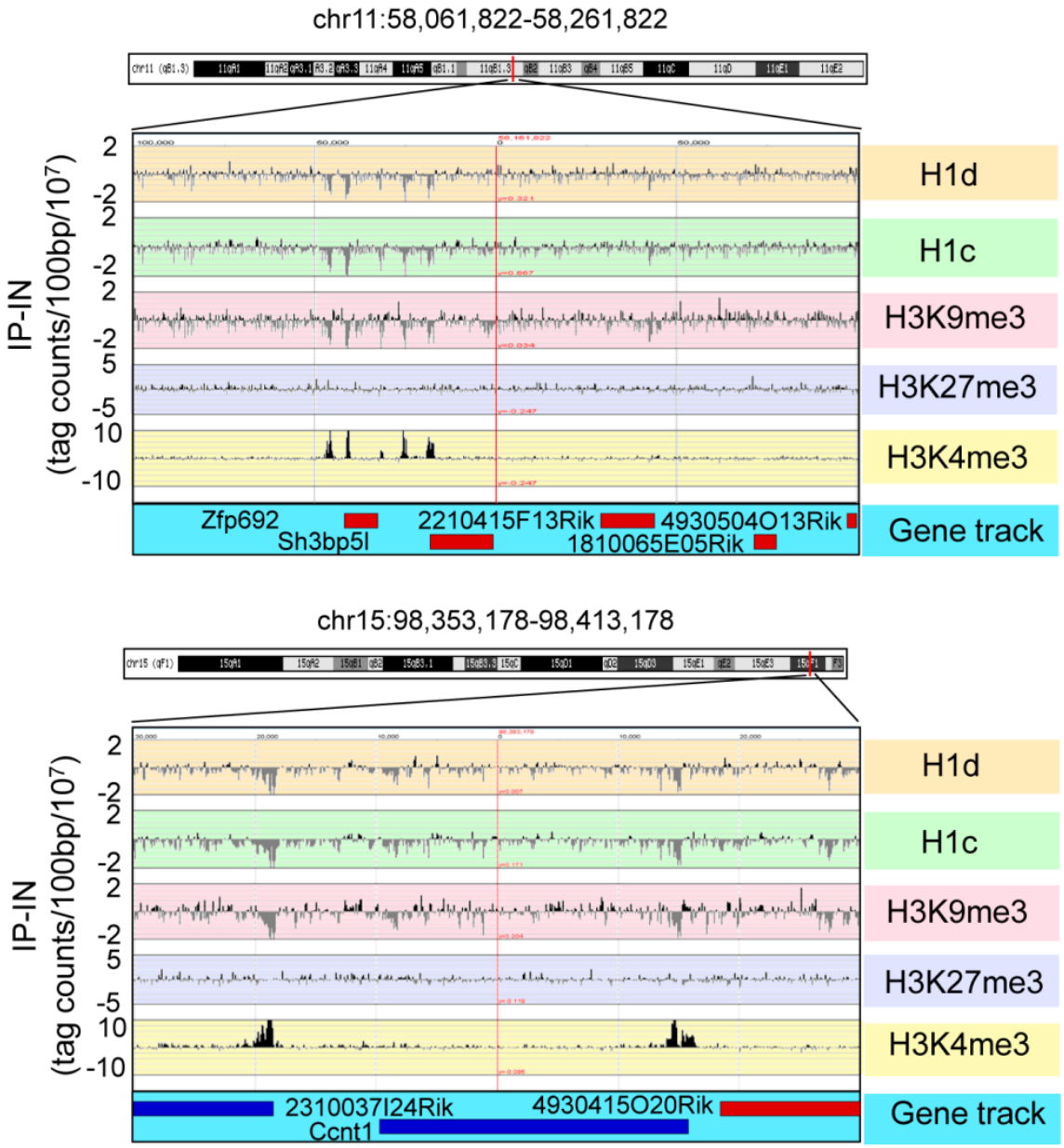

Figure 3.3 Examples of binding signals of H1d, H1c, and histone marks at TSSs.

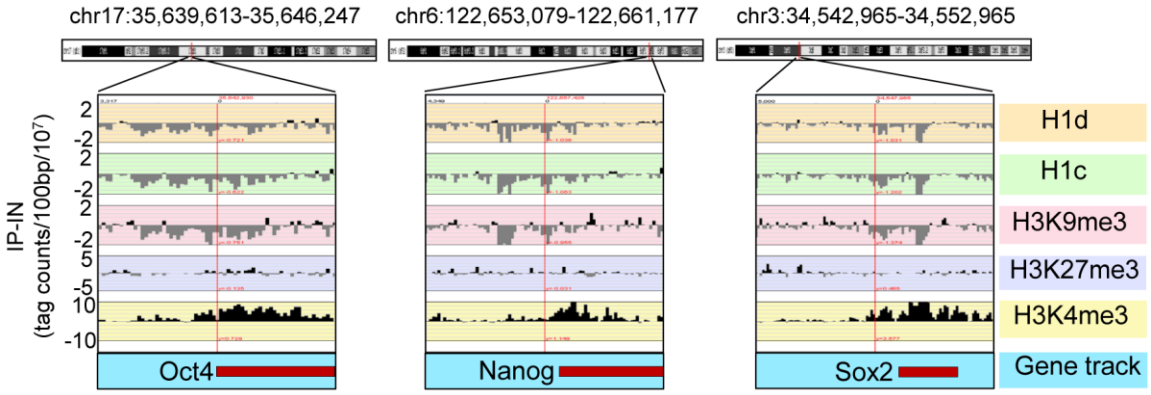

Figure 3.4. Distribution patterns of H1 variants and histone marks at C-Myc target genes.

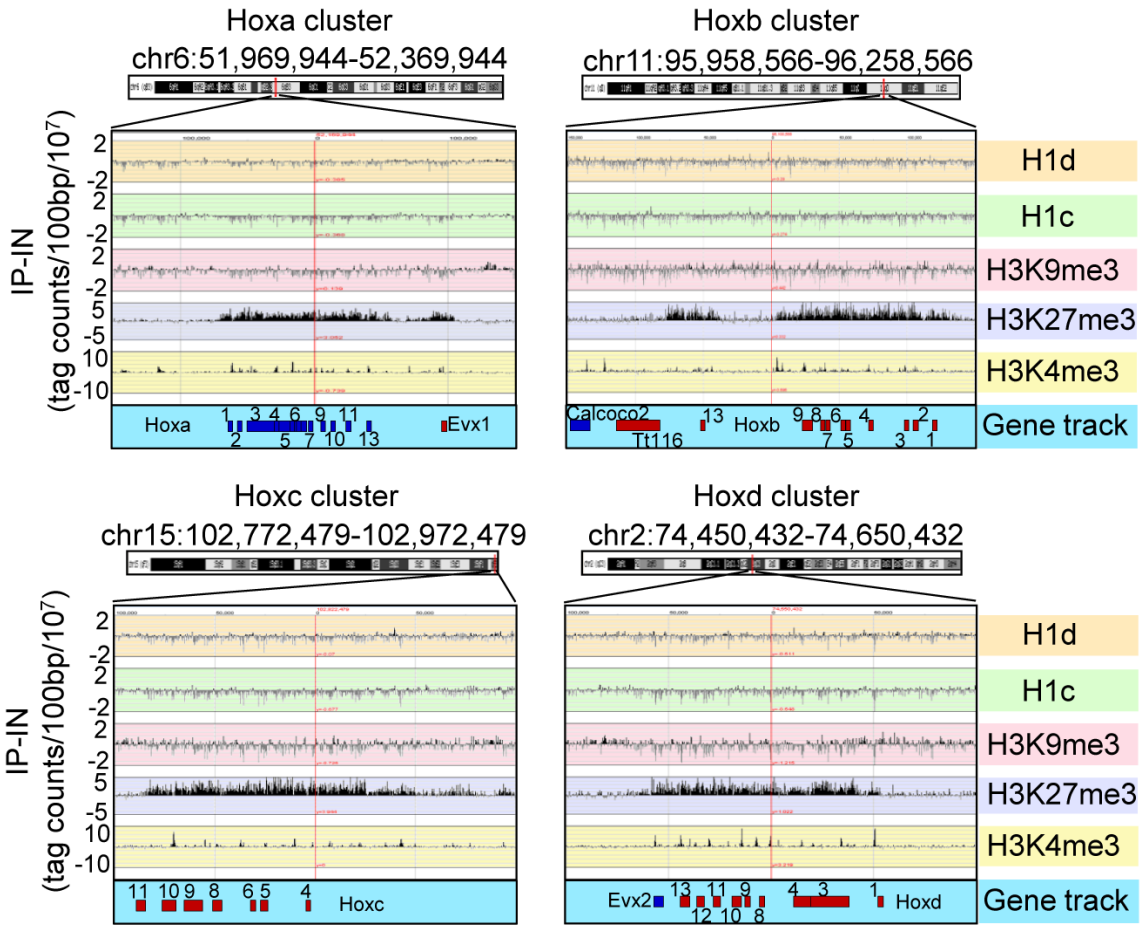

Figure 3.5 Occupancy of H1 variants and histone marks at 4 *Hox* clusters.

We next investigated the relationship between H1 occupancy and gene expression levels at a 10 kb region centered around transcription start sites (TSSs) as well as a 10 kb region centered around transcription termination sites (TTSs) using GenPlay software (Lajugie and Bouhassira 2011). Such metagene analysis revealed that H1 signals were always lower than chromatin input control within these regions (IP-IN<0) (Figure 3.6A), suggesting a general depletion of H1 at gene containing regions. Both H1d and H1c were especially depleted around the TSSs with dips much deeper at highly active genes than at silent genes (Figure 3.6A). Interestingly, except at TSSs and promoters, H1 signals remained largely constant throughout the gene encompassing regions and the signal intensity was higher at the silent genes than that at active genes, suggesting that H1 is underrepresented at surrounding regions of active genes as well (Figure 3.6A). Indeed, for genes highly depleted of H1 variants at promoters, the signal values of H1s, although gradually increased toward distal regions, remained diminished up to 200 kb from TSS (Figure 3.7), suggesting that H1s are depleted from broad domains at these regions in the genome. H3K4me3 is known to be peaked around TSS of active genes (Guenther et al. 2007; Mikkelsen et al. 2007), and metagene H3K4me3 curves displayed an opposite pattern to that of H1 (Figure 3.6A), further indicating that H1 is absent at active promoters. H3K9me3 exhibited a very similar distribution pattern to that of H1d and H1c, whereas H3K27me3 did not show similar profiles to that of H1 variants (Figure 3.6A). Metagene analysis of H1 and histone marks on genes finely partitioned by expression levels (each group with 20% of genes) over a 10 kb region (-5 kb to +5 kb of TSS) further corroborated their distinctive patterns at TSSs as a function of gene expression (Figure 3.6B).

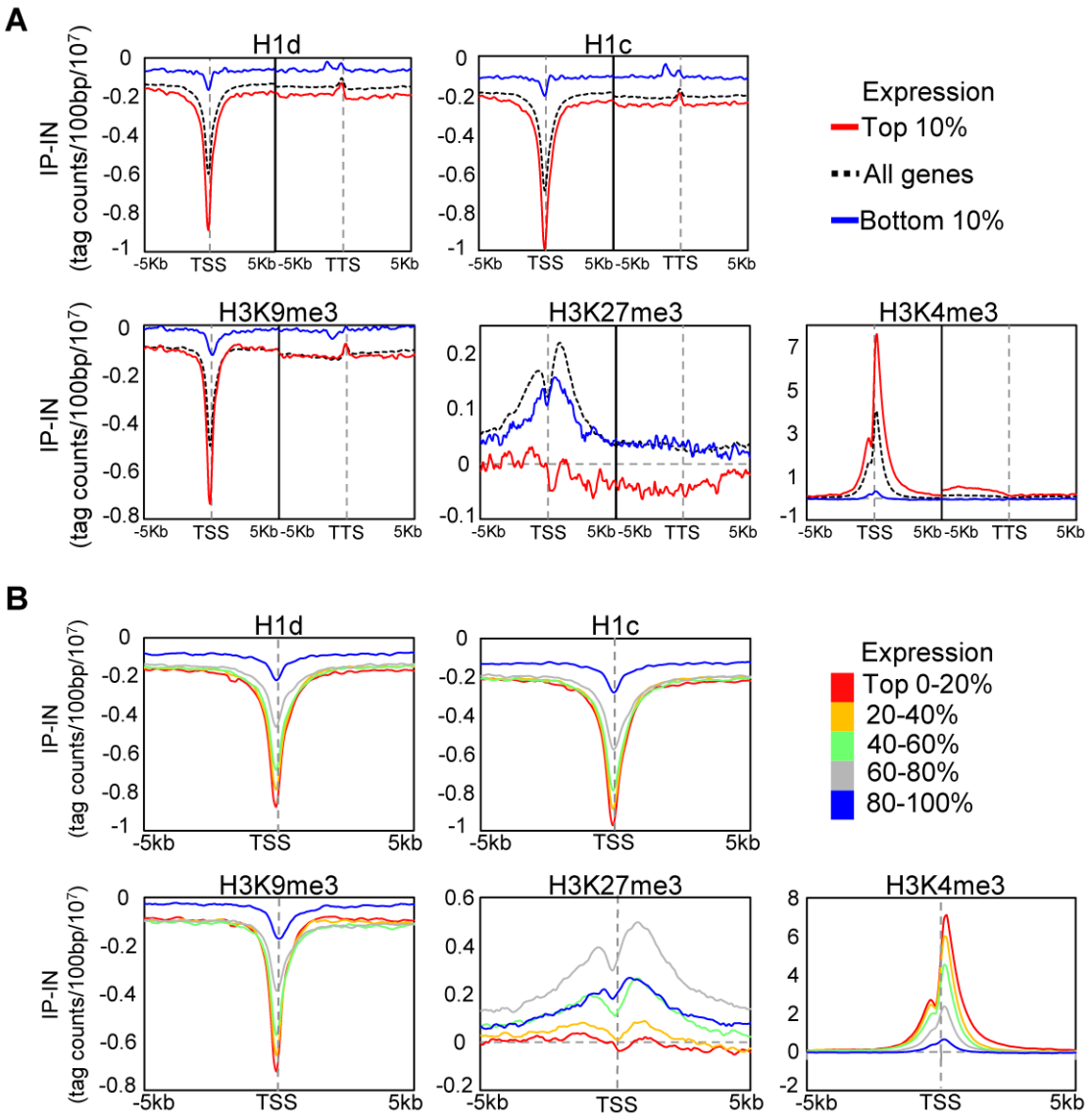

Figure 3.6 Metagene analysis of H1d, H1c, H3K9me3, H3K27me3 and H3K4me3 in relation to gene expression levels.

(A) Profiles of highly active genes (top 10% in expression), silent genes (bottom 10% in expression) and all genes on a 10 kb window around TSS and a 10 kb window around TTS.

(B) Profiles of genes finely grouped according to expression levels on a 10 kb window centered on TSSs. TSS: transcription start site. TTS: transcription termination site.

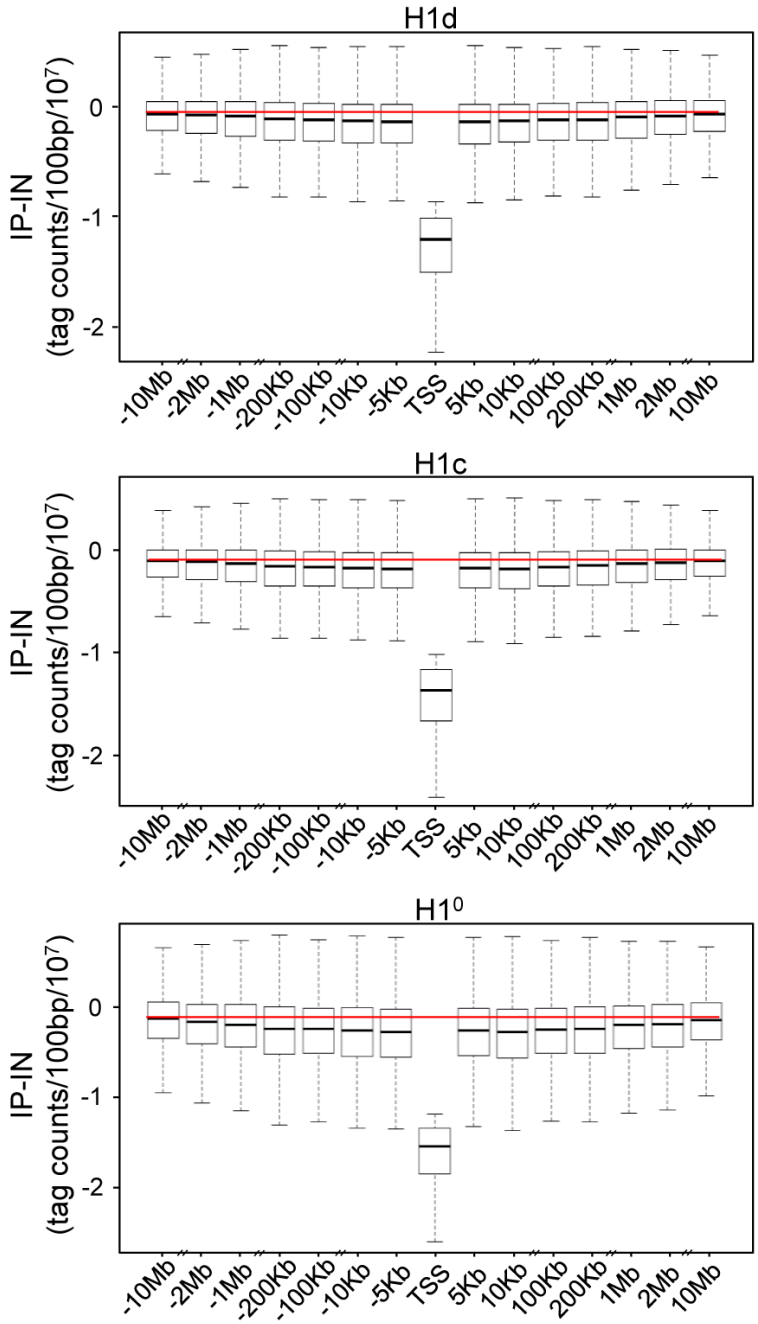

Figure 3.7 Progressively elevated levels of H1 variants with increasing distance from TSS.

Signal values of 100 bp window at TSS and indicated flanking regions of genes with lowest H1 values (20% of all genes) were plotted. Distal data points situated in the vicinity of other TSSs were removed from calculation. $P<10^{-50}$ for all comparisons (with TSS) with paired t-test. The line in the box indicates the median, while the bottom and top of the boxes are the 25th and 75th percentiles, respectively. The red line represents the median signals at +/-10 Mb distal to TSS.

To better define the correlation of H1 occupancy with histone marks around TSSs and promoters, metagene analysis of H1 signals was performed for genes partitioned into 5 groups according to their levels of H3K9me3, H3K4me3, or H3K27me3, which displayed characteristic profiles around TSS (Figure 3.9A, 3.9B, 3.9C and Figure 3.8). H1 signals displayed positive and negative correlations with respective H3K9me3 and H3K4me3 signals, having the deepest dip for promoters and TSSs with the lowest H3K9me3 levels (Figure 3.9A) or highest H3K4me3 signals (Figure 3.9B). On the other hand, H1 signals showed no correlation with H3K27me3 levels and no difference among the 5 groups of genes partitioned according to H3K27me3 levels (Figure 3.9C). Interestingly, H1 was also depleted at the promoters of genes bound by H3K4me3 and H3K27me3 bivalent marks (Bernstein et al. 2006) but not at H3K4me3-free promoters, regardless of the presence or absence of H3K27me3 (Figure 3.9D).

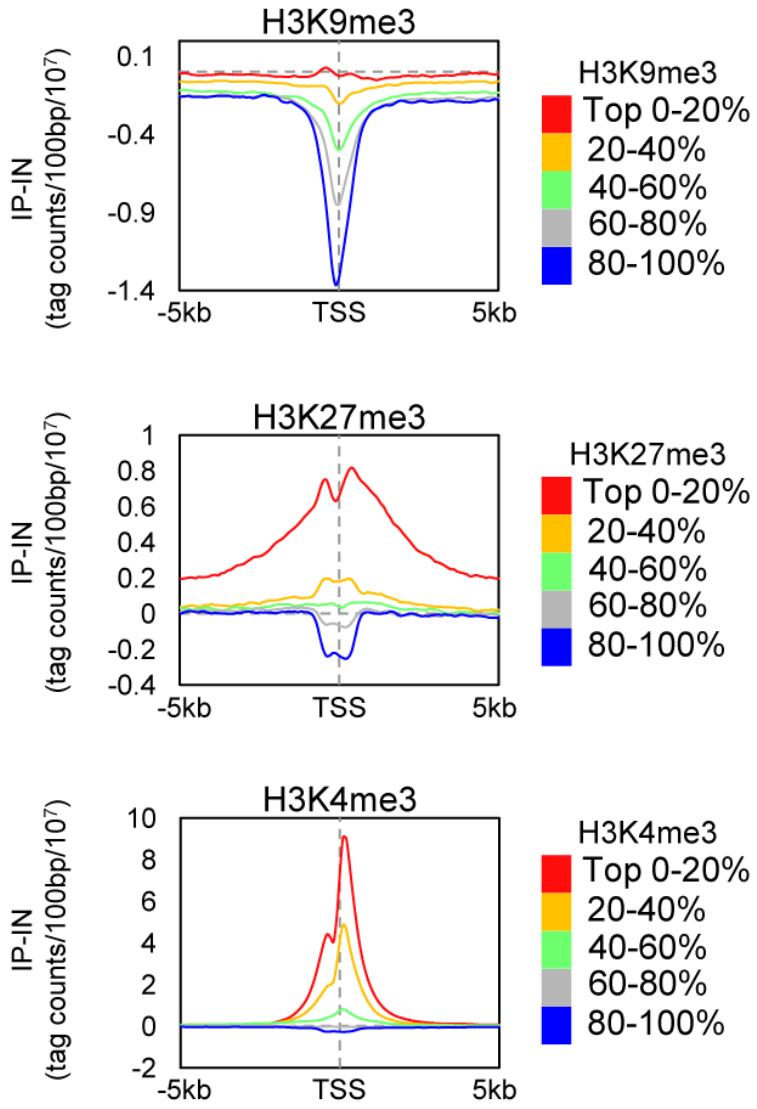

Figure 3.8 Metagene profiling analysis of H3K9me3, H3K27me3, and H3K4me3 around TSS in relation to levels of themselves.

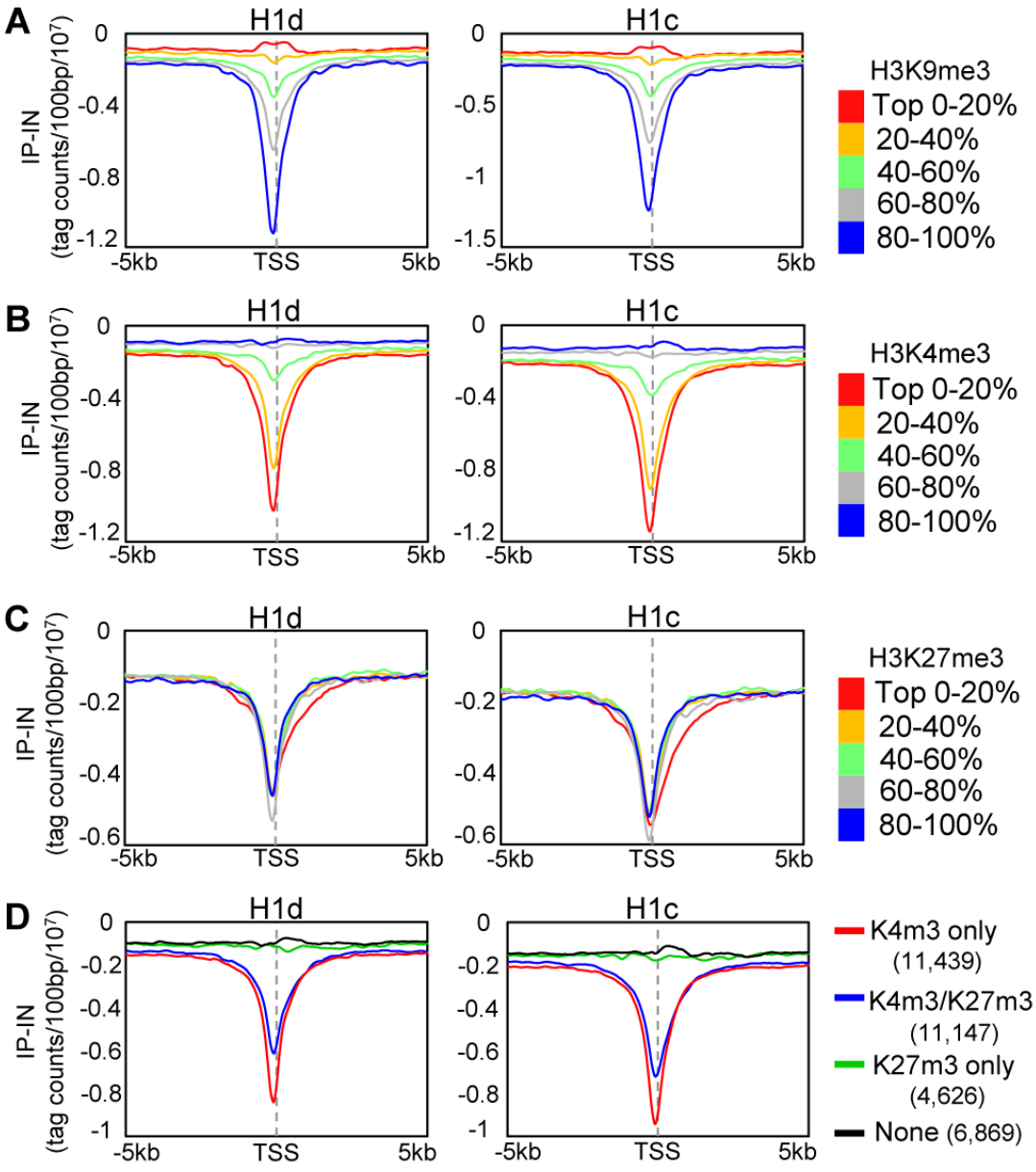

Figure 3.9 Metagene analysis of H1d and H1c in relation to the levels of H3K9me3 (A), H3K4me3 (B), H3K27me3 (C), and the presence or absence of H3K4me3 and H3K27me3 (D), on regions covering -5 kb to + 5 kb of TSS.

The number of genes selected within each group in (D) is shown in parentheses. Y axis: tag counts per 100 bp window per 10 million mappable reads. IP-IN: normalized signal values of ChIP-seq subtracted by that of input-seq.

Although most H1d and H1c signals appeared universally distributed, we identified regions enriched for H1 binding using SICER (Zang et al. 2009) and GenPlay software. Identified H1d and H1c enriched regions often form broad domains (examples shown in Figure 3.10). SICER analyses were performed through collaboration with Karan Uppal from Dr. Eva Lee's lab as well as Fan lab members Ashwath Kumar and Xiao Dong. Annotation of H1d- and H1c- rich regions using CEAS (Shin et al. 2009), a software designed to characterize both sharp and broad ChIP-seq enrichment, indicated that, similar to H3K9me3, both H1d and H1c "peaks" were over-represented in distal intergenic regions and under-represented at promoters and 5'UTR, which were highly enriched with H3K4me3 peaks as reported previously (Figure 3.11 and (Mikkelsen et al. 2007)).

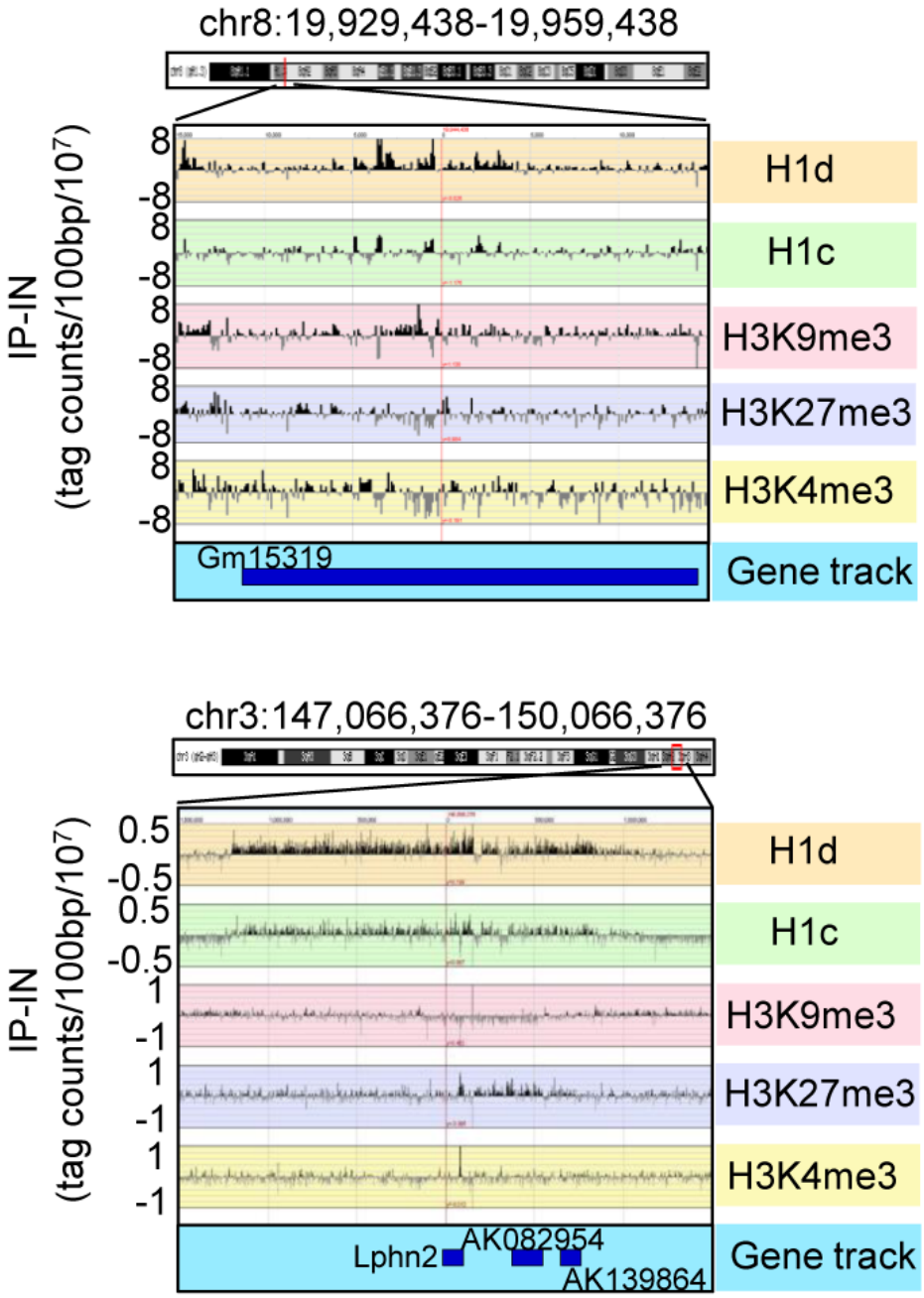

Figure 3.10 Examples of H1d and H1c enriched regions.

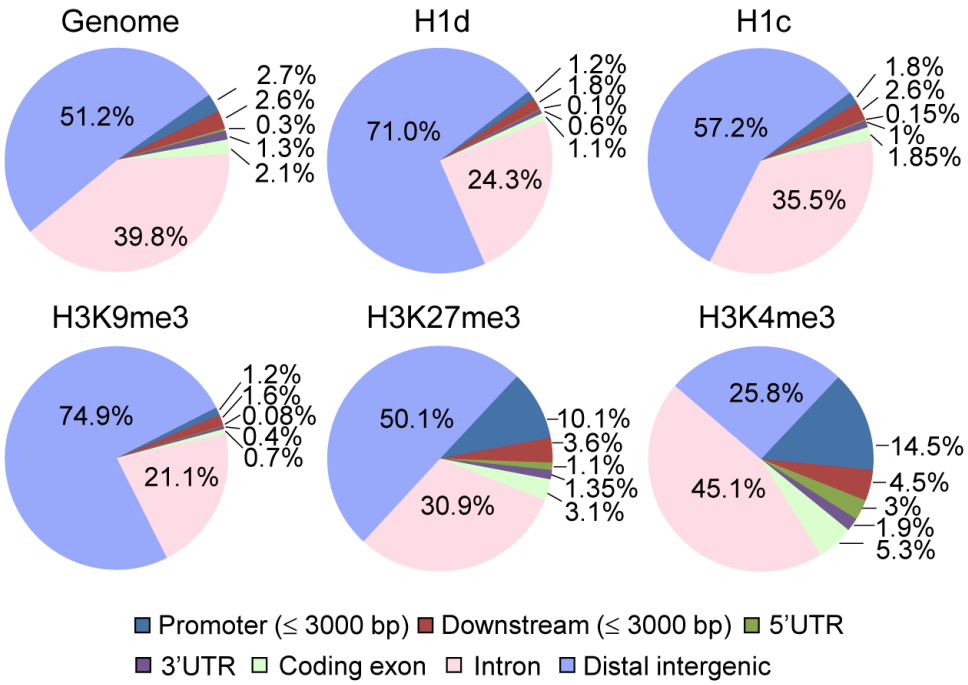

Figure 3.11 Distribution analysis of H1d and H1c enriched regions.
Pie diagram of distributions of H1d, H1c, H3K9me3, H3K27me3, and H3K4me3 enriched regions at genes, proximal regulatory regions, and distal intergenic regions.

3.4.2 Correlation of H1 with histone marks

We next performed genome-wide correlation analysis to determine if the similarity and/or contrast of H1 variants with GC% and histone marks at TSSs also extend to a genome-wide scale. Indeed, the distribution of H1d and H1c were highly correlated throughout the genome (R= 0.7866) (Figure 3.12), and both variants were negatively correlated with GC% (R= -0.4182 and -0.4140 for respective H1d and H1c) (Figure 3.13), indicating that H1d and H1c were enriched or depleted at similar regions. Both H1d and H1c were correlated negatively with H3K4me3 (R= -0.2640 and -0.3317 respectively), but positively with H3K9me3 (R= 0.5732, 0.5790) (Figure 3.13), suggesting their enrichment at heterochromatin. On the other hand, these two variants showed no obvious correlation with H3K27me3 (R= -0.08 for both variants) (Figure 3.13). Correlation analysis of sequencing signals on enriched or depleted regions gave similar coefficients as the respective genome-wide coefficients (data not shown). It is interesting to note that the coefficients of H1 *vs.* H3K4me3 on sex chromosomes were dramatically different from those of autosomes (Figure 3.14). This result echoes the previous finding that sex-chromosome genes are overrepresented among genes with altered expression levels by triple H1 deletion in ESCs (Fan et al. 2005), suggesting that H1 may play a role in regulating higher order chromatin structures of sex chromosomes.

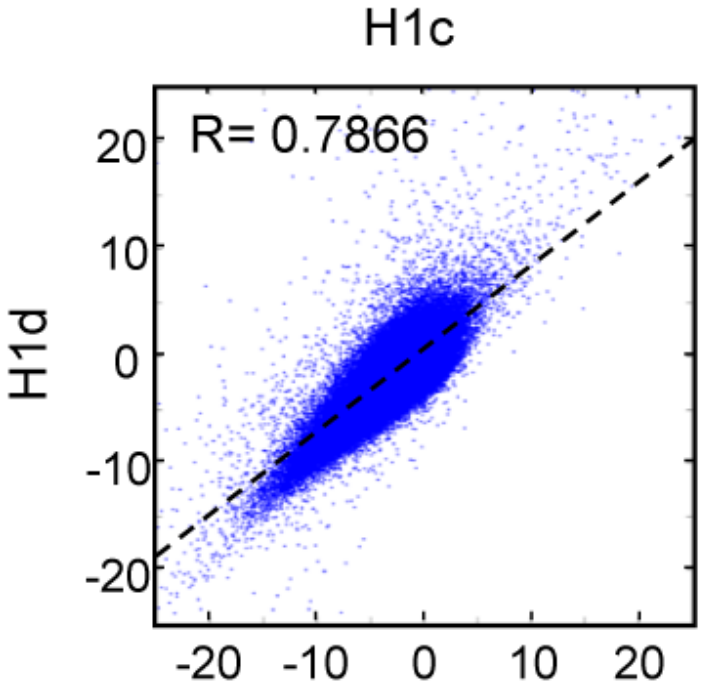

Figure 3.12. Genome-wide correlation scatter plots of H1d *vs.* H1c.
The correlation coefficient and the trend line were generated as described in Materials and Methods. X and Y axes: average signal values (normalized to 100 bp window). Pearson's correlation was used to perform the analysis. $P<10^{-100}$ for the correlation coefficient.

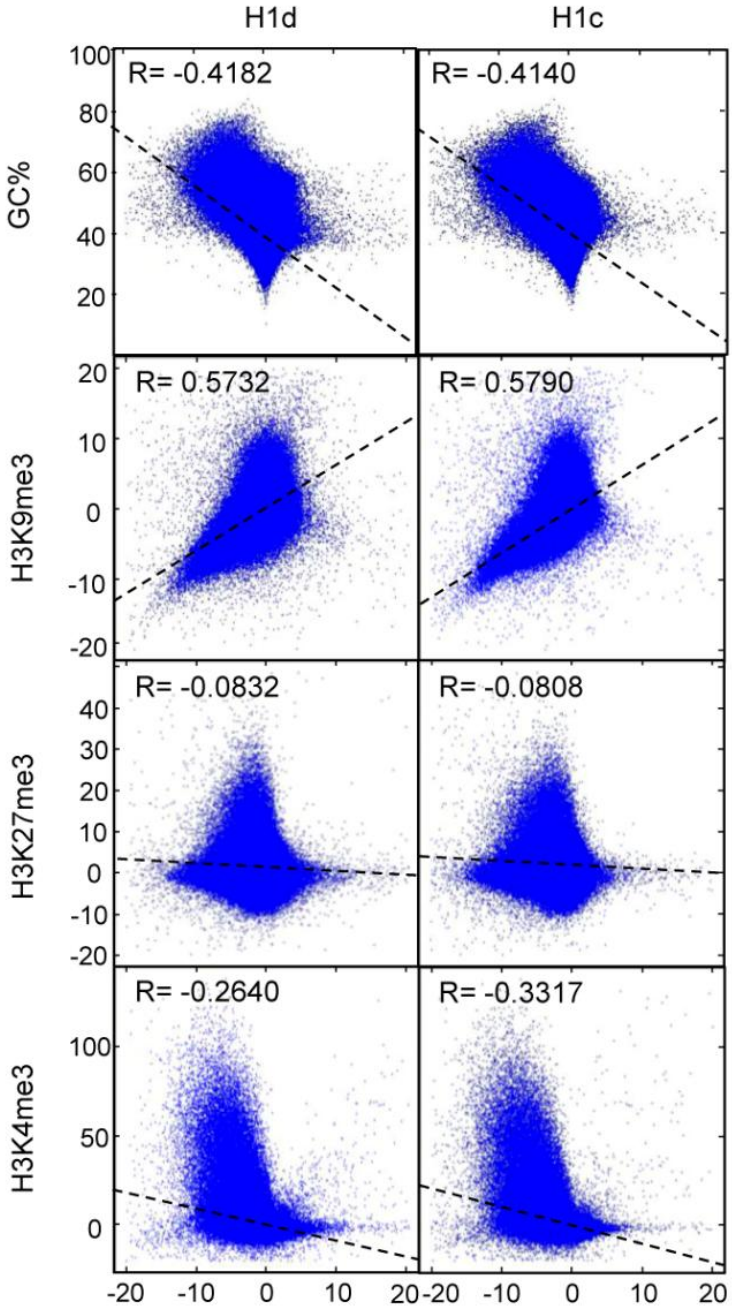

Figure 3.13 Genome-wide correlation scatter plots of GC% (or histone marks) *vs.* H1d (left) and H1c (right).

The correlation coefficients and the trend lines were generated as described in Materials and Methods. X and Y axes: average signal values (normalized to 100 bp window). Pearson's correlation was used to perform the analysis. $P<10^{-100}$ for all correlation coefficients.

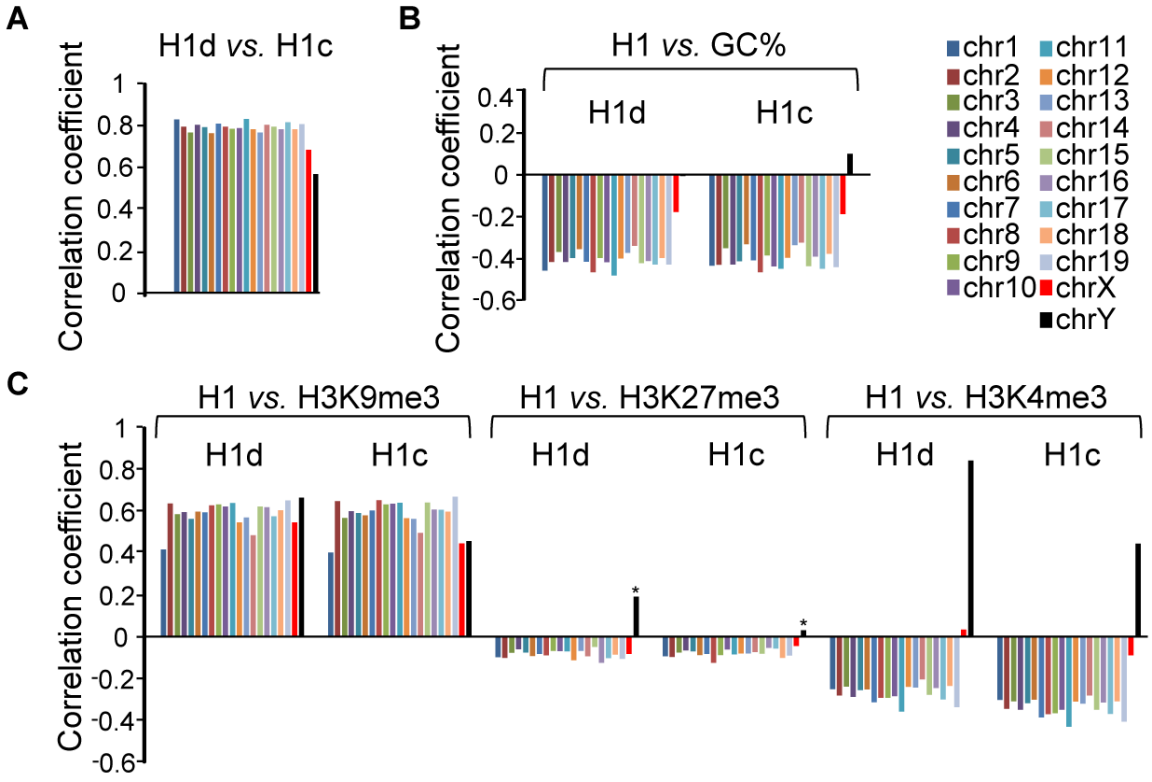

Figure 3.14 Correlation coefficients of H1d *vs.* H1c (A), each H1 variant (H1d or H1c) *vs.* GC percentage (B) and histone marks (C) on individual chromosomes.

Pearson's correlation was used to perform the analysis. $P<10^{-100}$ for all correlation coefficients except for those labeled with "*". *: P>0.001.

3.4.3 Features of H1d and H1c enriched regions

To gain a comprehensive view of the DNA features of H1d- and H1c- rich regions, I collaborated with Ashwath Kumar, who selected the regions highly enriched for H1 variants and histone marks, and performed cross-comparison of genome attributes using the statistical analysis software EpiGRAPH (Bock et al. 2009). Such analysis (Figure 3.15 and Figure 3.16) revealed that: *a*) H1d/H1c common peaks (regions highly enriched for both H1d and H1c) appeared similar to H3K9me3 peaks in genome attributes, except for satellite DNA which was relatively overrepresented in H1 peak regions; *b*) H1d/H1c common peaks were enriched at AT-rich sequences, satellite DNA, and chromosome G-bands but were absent from GC-rich regions, and genes or exons when compared with H3K4me3 or H3K27me3 peaks; *c*) comparison of H1d/H1c common peaks with H1d/H1c unique peaks (regions highly enriched for H1d or H1c but not both) showed similar features as the comparison of H1d/H1c common peaks with H3K4me3 or H3K27me3 peaks; *d*) comparison of H1d *vs.* H1c specific peaks indicated that H1d unique peaks were relatively enriched at GC-rich sequences and LINEs, whereas H1c unique peaks were more enriched at AT-rich sequences, Giemsa positive regions and satellite DNA; *e*) the overrepresentation analyses between H1d (or H1c) unique peak regions and histone mark peak regions exhibited similar features as comparisons using H1 common peaks. These results define common and unique features for H1d and H1c enriched regions.

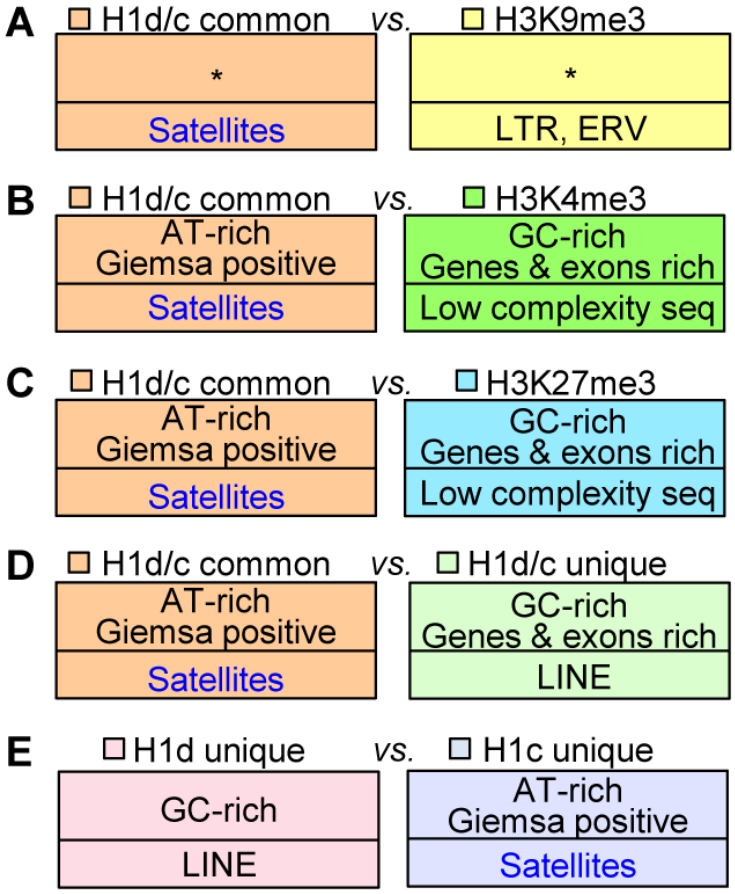

Figure 3.15 Overrepresented features from 5 comparisons of H1 or histone mark highly enriched regions.
(A) H1d/H1c common *vs.* H3K9me3 regions; (B) H1d/H1c common *vs.* H3K4me3 regions; (C) H1d/H1c common *vs.* H3K27me3 regions; (D) H1d/H1c common *vs.* H1d/H1c unique regions; (E) H1d unique *vs.* H1c unique regions.
Bottom half of each box: repetitive elements. *: no significant overrepresentation. All P values remained significant after multiple testing corrections with the FDR method and the more conservative Bonferroni method.

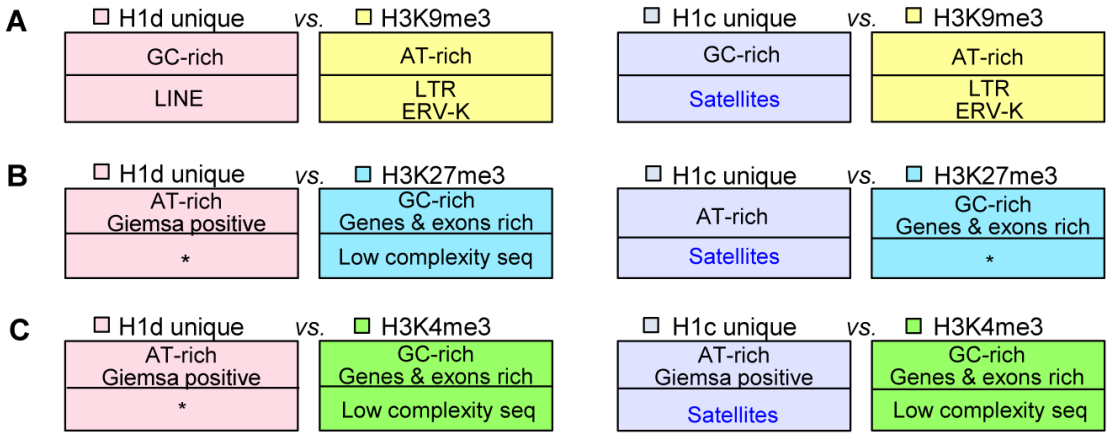

Figure 3.16 EpiGRAPH overrepresentation analyses of comparisons of H1d (or H1c) uniquely enriched regions *vs.* histone marks enriched regions.
H1d unique regions (left panels) or (H1c unique regions (right panels)) *vs.* H3K9me3 regions (A), *vs.* H3K27me3 regions (B), *vs.* H3K4me3 regions (C). Overrepresented repetitive elements are shown in the bottom half of each box. *: no significantly overrepresented features. All P values remained significant after multiple testing corrections with the FDR method and the more conservative Bonferroni method.

3.4.4 High occupancy of H1d and H1c at major satellite sequences in mESCs

The EpiGRAPH overrepresentation analysis indicated that peak regions of H1d and H1c were enriched for satellite repeats. Indeed, examination of the top ranked H1 peak regions with especially high binding signals revealed that these regions overlap perfectly with major satellite sequences (examples shown in Figure 3.17). This finding and the above observation of overrepresentation of multi-match sequence reads in H1 ChIP-seq libraries prompted us to perform a thorough mapping study of sequence reads to a database of repetitive sequences. We aligned sequence reads of H1d, H1c, H3K9me3, H3K27me3 and H3K4me3 ChIP-seq libraries to Repbase Update, a comprehensive database of repetitive elements from diverse eukaryotic organisms (Jurka 2000; Jurka et al. 2005; Bock et al. 2010). We found that both H1d and H1c were significantly enriched at repetitive sequences, with H1d and H1c ChIP-seq libraries having on average percent mapped repeats respective 2.3-, and 2.8-fold of that of chromatin input-seq libraries (Figure 3.18). H3K9me3, H3K27me3 and H3K4me3 ChIP-seq libraries had an average respective percent mapped repeats 1.4-, 0.7-, and 0.9- fold compared with input controls (Figure 3.18), suggesting an overrepresentation of H3K9me3, yet not as dramatic as H1d and H1c, at repetitive sequences.

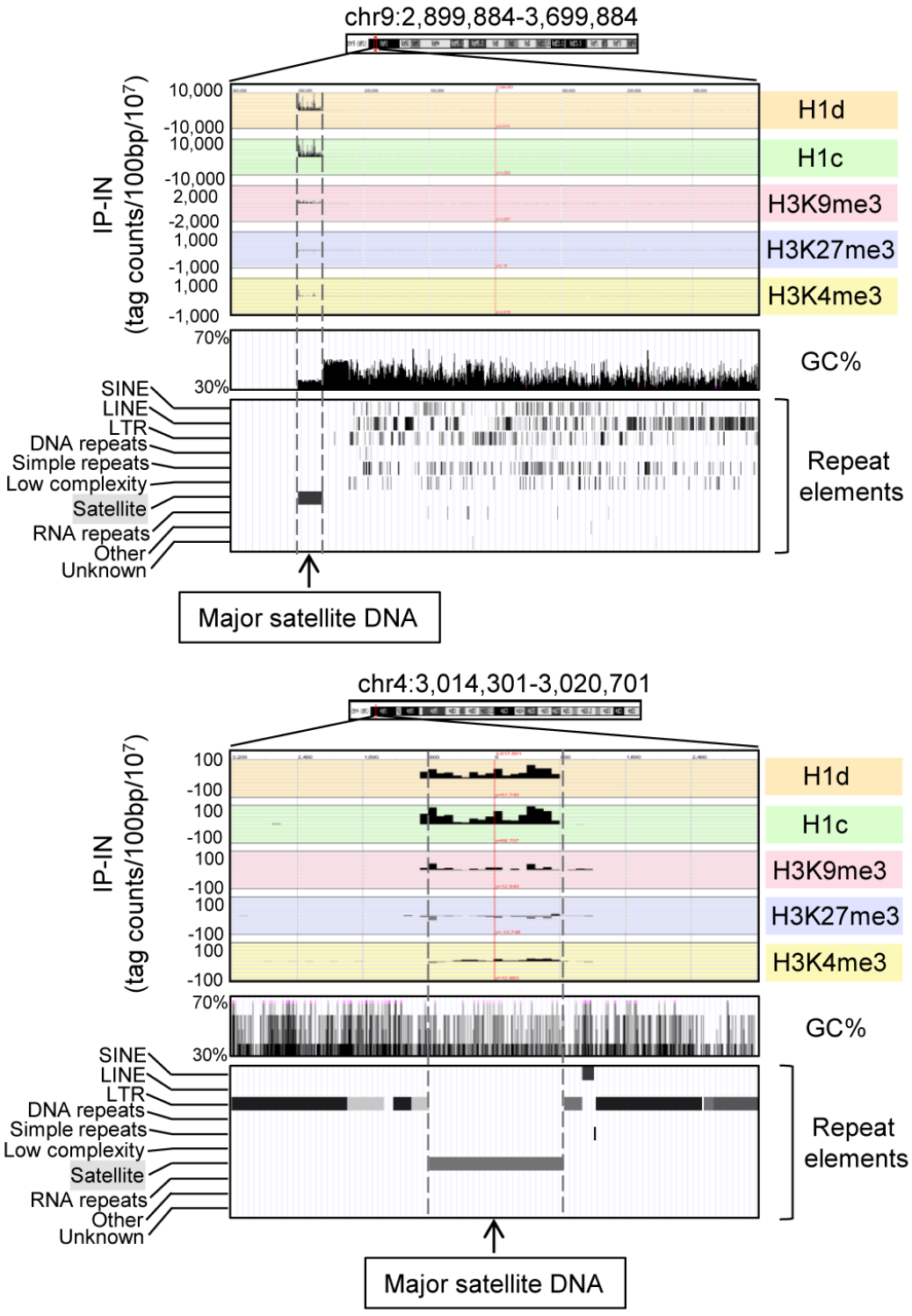

Figure 3.17 Representative profiles of top H1d and H1c enriched regions (mapped to mm9).

Repeat element tracks were obtained from UCSC genome browser. Dashed lines indicate the localization of these H1 peaks at major satellite sequences.

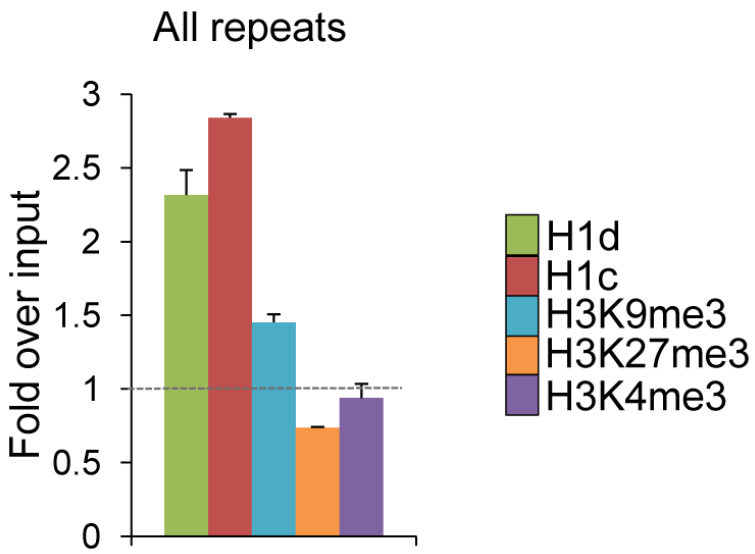

Figure 3.18. H1d and H1c are enriched at repetitive sequences.
Fold enrichment of percent mappable repeats (mapped to RepBase) from H1d, H1c, and histone marks ChIP-seq libraries over that from corresponding chromatin input-seq library on all repeats. The dashed lines indicate the level of normalized input signal. P values calculated with Fisher's exact test comparing ChIP-seq with input-seq libraries are less than 2.5×10^{-5}. Error bars represent the differences between replicates. Data are presented as average ± S.E.M.

Importantly, we found that the increased proportion of total reads of H1 libraries mapped to repetitive sequences was predominantly caused by overrepresentation on the major satellite sequences on which the levels of H1d and H1c occupancy were enriched on average 4.0- and 5.6-fold compared with the chromatin input control (Figure 3.19). This level of H1 enrichment appeared to be specific to major satellites because we did not observe H1d and H1c enrichment among other abundant repeats, except for a moderate increase of H1d and H1c occupancy at minor satellites. qChIP results confirmed the preferential binding of these two H1 variants to major satellites (Figure 3.20). Sequencing results showed that H1d and H1c levels on most of other less abundant classes of repetitive elements, such as L1, IAP LTR retrotransposons, SINE, non-LTR retrotransposons, and DNA transposons, were similar or lower compared with the input control (Figure 3.19 and Figure 3.21). H3K4me3 was highly enriched at 5'end of a subset of LINE L1 sequence (Figure 3.21), consistent with the abundant expression of L1 detected in multiple cell types (Packer et al. 1993; Trelogan and Martin 1995; Belancio et al. 2010), whereas H3K9me3 was enriched at major satellite repeats and LTR transposons, such as IAP particles, with similar levels as previously reported (Martens et al. 2005; Mikkelsen et al. 2007) (Figure 3.19 and Figure 3.21).

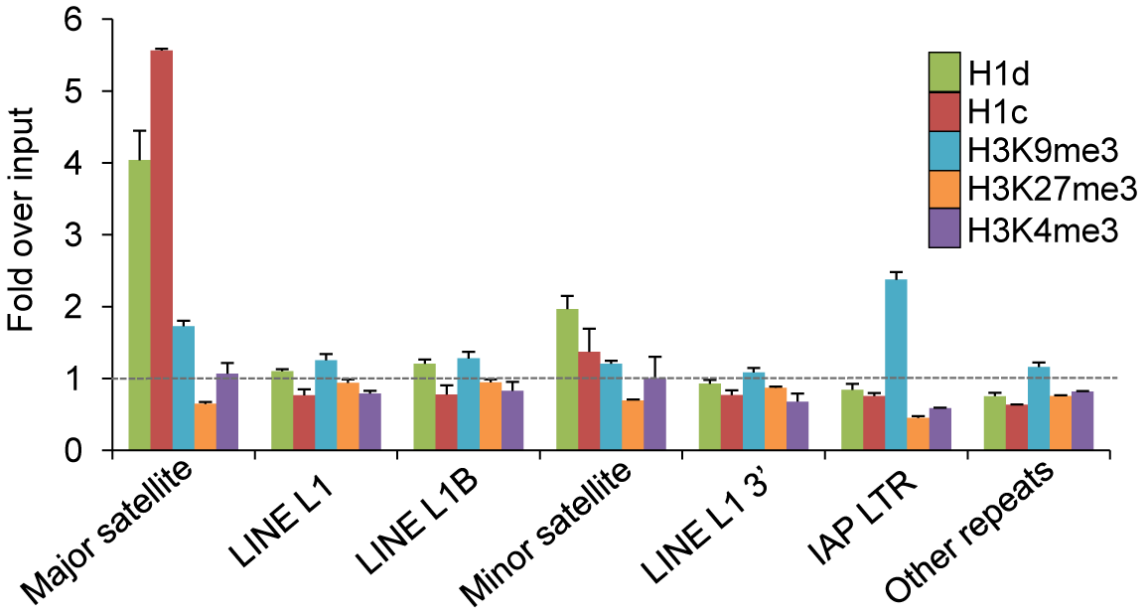

Figure 3.19 H1d and H1c are highly enriched at the major satellite sequences.
Fold enrichment of percent mappable repeats (mapped to RepBase) from H1d, H1c, and histone marks ChIP-seq libraries over that from corresponding chromatin input-seq library on six most abundant repetitive sequences and the remaining other repeats. The dashed lines indicate the level of normalized input signal. P values calculated with Fisher's exact test comparing ChIP-seq with input-seq libraries are less than 2.5×10^{-5} for all repeat classes shown. Error bars represent the differences between replicates. Data are presented as average ± S.E.M.

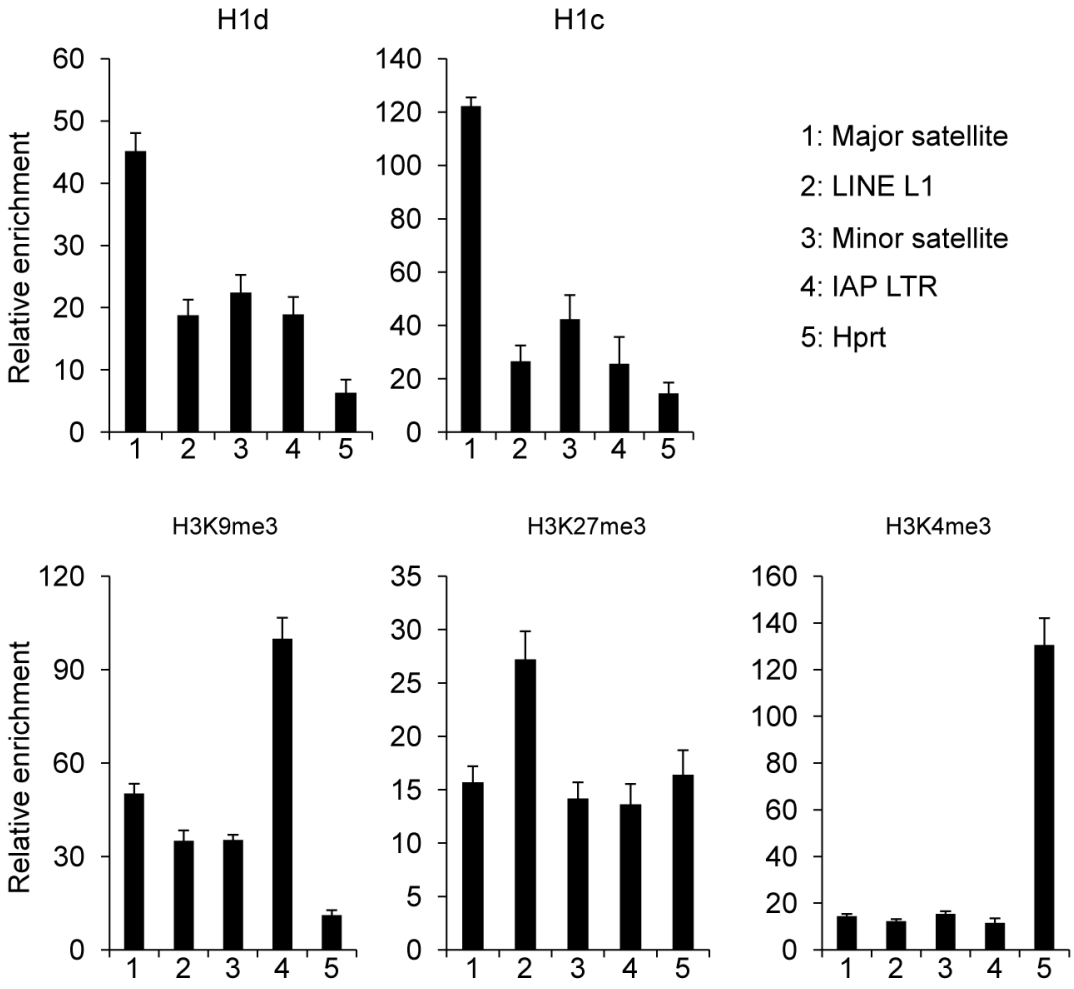

Figure 3.20 qChIP analysis of H1d, H1c and histone marks at selected repetitive elements.

Relative enrichment was calculated by normalizing the signals of ChIP over that of IgG. Data are presented as mean ± S.D.

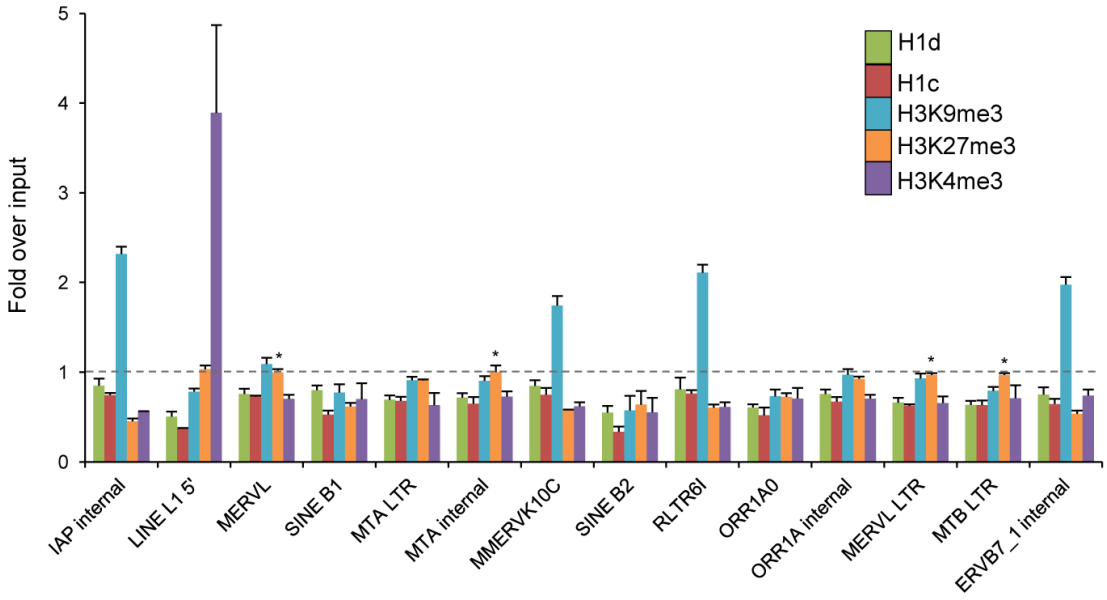

Figure 3.21 Distribution of H1d, H1c, and histone marks on additional repetitive sequences.

Fold enrichment of percent mapped repeats of H1d, H1c, H3K9me3, H3K27me3, and H3K4me3 ChIP-seq libraries over that of corresponding input-seq library. 14 most abundant repetitive sequences within the "other" repetitive group shown in Figure 3.19 are presented. P values calculated with Fisher's exact test comparing ChIP-seq with input-seq libraries are less than 1.3×10^{-7} for all repeat classes shown except those marked with "*". *: P>0.01. Error bars represent the differences between replicates. Data are presented as average ± S.E.M.

Enrichment of H1 variants at major satellites was also confirmed by calculating the normalized "IP-IN" signals at major satellite regions in mouse genome mm9 assembly (July 2007) annotated by RepeatMasker (http://repeatmasker.org) (Figure 3.22 and Figure 3.23). Analysis of ChIP-seq libraries of FLAG-H1d in H1d-*trans* ESCs, which had similar levels of FLAG-H1d and total H1/nuc ratio as the *cis* $H1d^{FLAG}$ ESCs (Figure 3.24), also showed a similar level of enrichment at major satellites as $H1d^{FLAG}$ ESCs (Figure 3.25, Figure 3.26, and Figure 3.27).

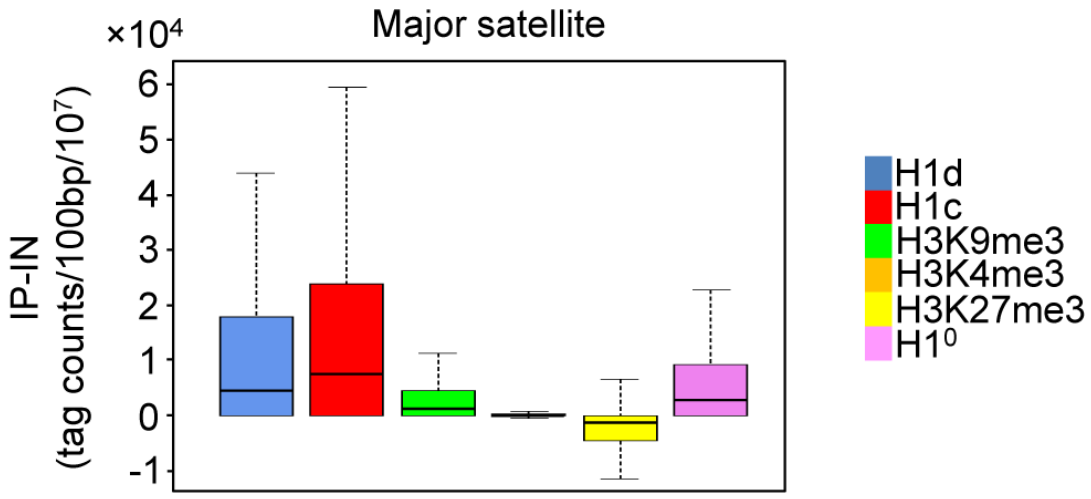

Figure 3.22 Significant enrichment of H1 variants at major satellites.
Box plots of the signals of three H1 variants and histone marks at major satellite repeats are shown.
Y axis: input subtracted, normalized signal values as tag counts per 100 bp window per 10 million mappable reads. The line in the box indicates the median, while the bottom and top of the boxes are the 25th and 75th percentiles, respectively. $P<4\times10^{-6}$ for comparisons with H3K9me3 by unpaired t-tests.

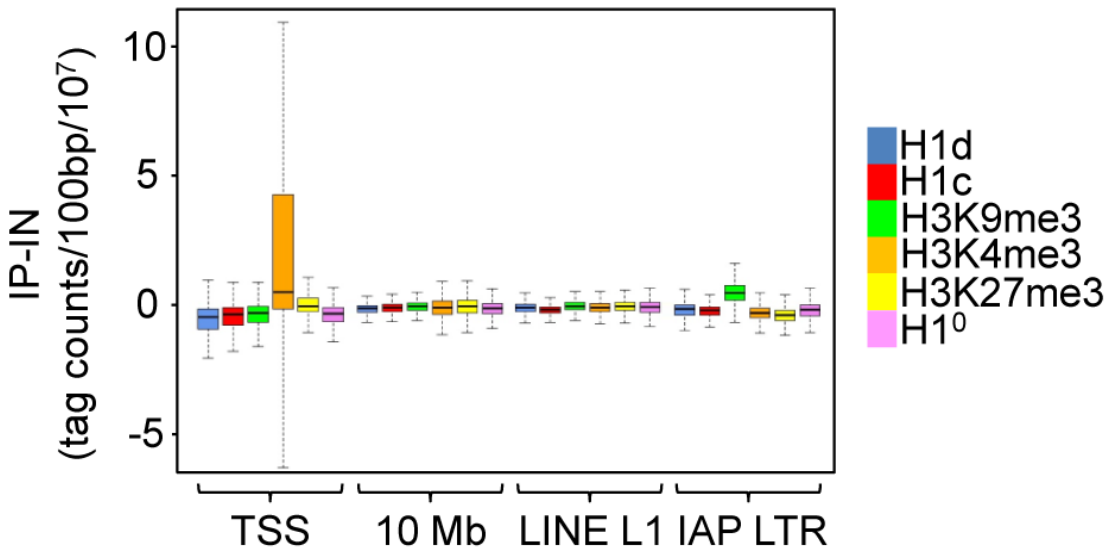

Figure 3.23 Binding signals of H1 variants and histone marks at TSS, 10 Mb distal to TSS, LINE L1, and IAP LTR repeats.
Y axis: input subtracted, normalized signal values as tag counts per 100 bp window per 10 million mappable reads. The line in the box indicates the median, while the bottom and top of the boxes are the 25^{th} and 75^{th} percentiles, respectively. $P<4\times10^{-6}$ for all comparisons with H3K9me3 within each category by unpaired t-tests.

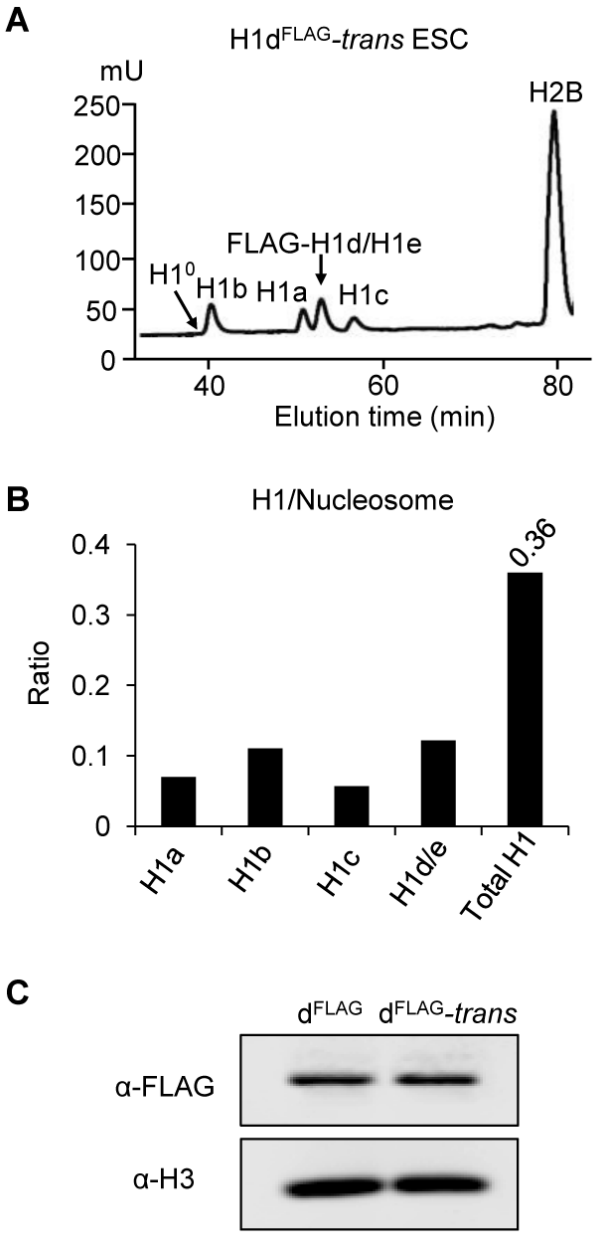

Figure 3.24 Analysis of H1d-*trans* ESC line.
(A) Reverse phase HPLC of total histone extracts from H1d-*trans* cells.
(B) Ratios of each H1 variant (and total H1) to nucleosome of H1d-*trans* cells calculated from data shown in (A).
(C) Western blotting indicating similar amount of FLAG-H1d in H1d^{FLAG} and H1d-*trans* cells.

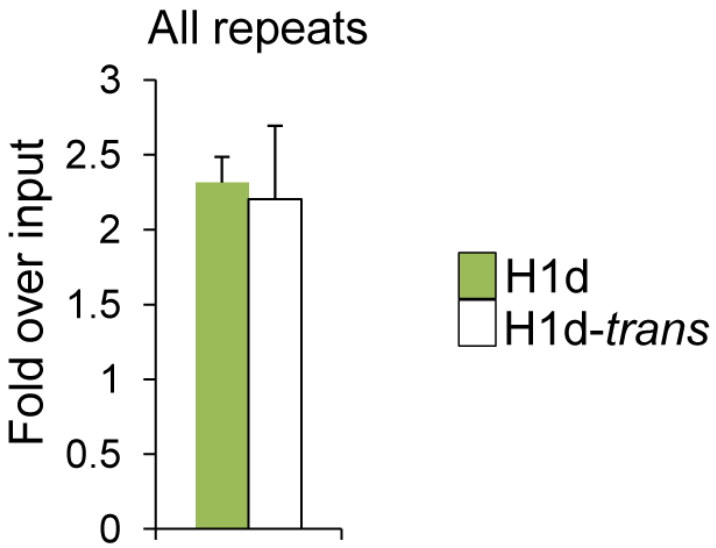

Figure 3.25 Similar enrichment of H1d at repetitive sequences in H1d^{FLAG} and H1d-*trans* ESCs.

P values comparing ChIP-seq with input-seq libraries are less than 9.3×10^{-14}.

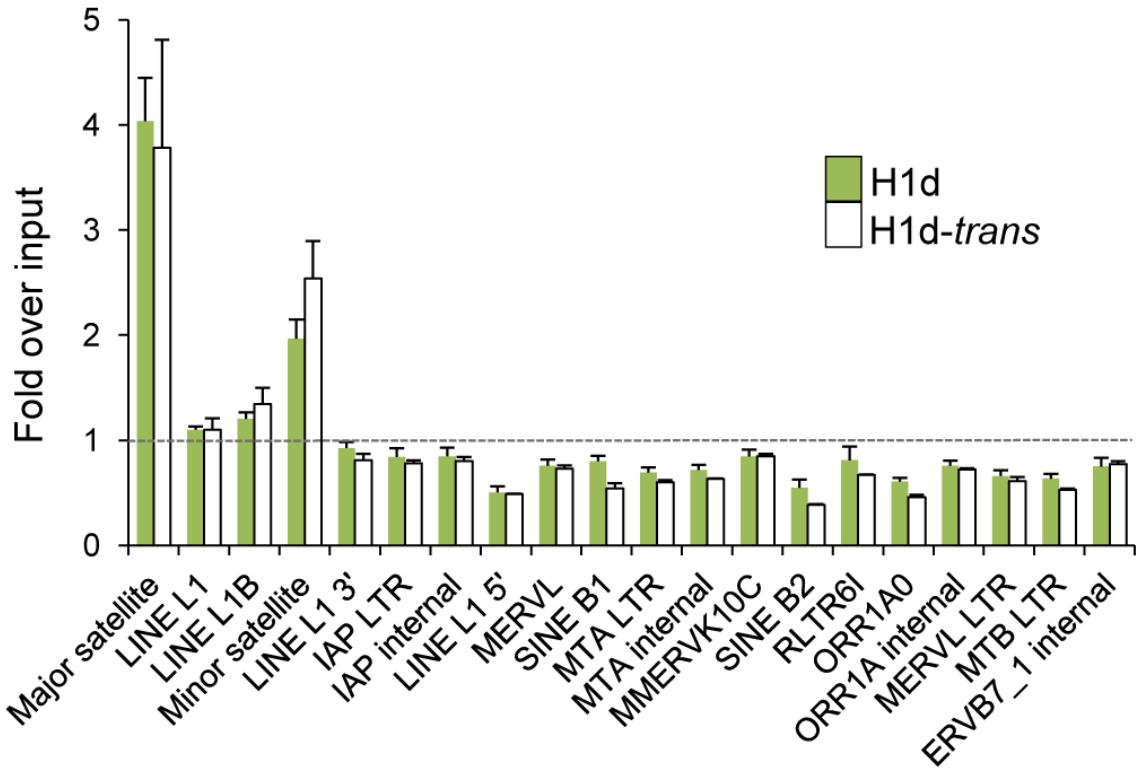

Figure 3.26 Similar binding of H1d at the 20 most abundant repetitive elements in H1d^{FLAG} and H1d-*trans* ESCs.

P values comparing ChIP-seq with input-seq libraries are less than 9.3×10^{-14} for all repeat classes shown.

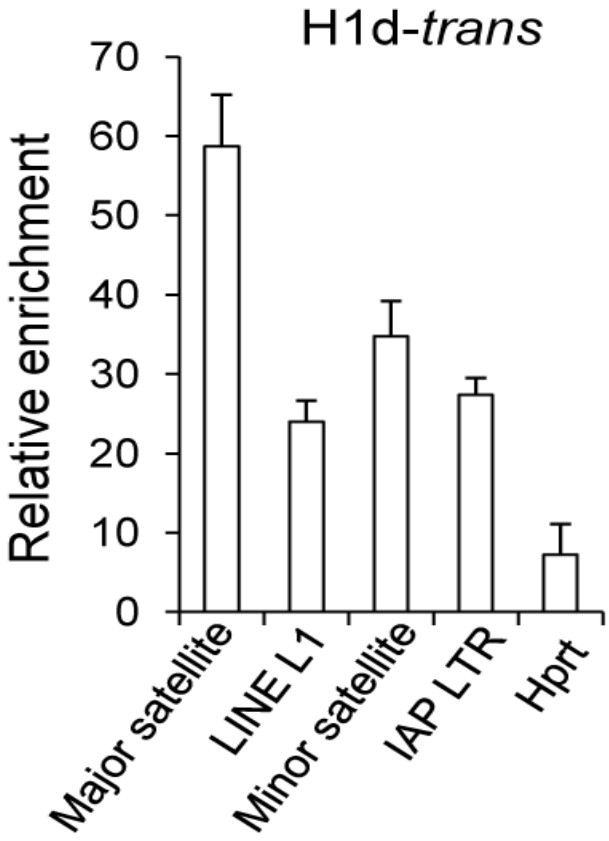

Figure 3.27 qChIP analysis of H1d occupancy at indicated repetitive elements in H1d-*trans* cells.
Relative enrichment was calculated by normalizing the signals of ChIP over that of IgG.

3.4.5 Mapping of $H1^0$ in mESC genome

To compare the binding patterns of the replacement variant $H1^0$ with that of the somatic variants H1d and H1c in ESCs, we performed ChIP-seq of $H1^0$ in $fH1^0$ cells with an anti-FLAG antibody. Such analysis indicated that, despite its different biochemical properties and unique expression patterns (Lennox and Cohen 1983; Zlatanova and Doenecke 1994; Happel and Doenecke 2009), $H1^0$ shared similar distribution features to that of H1d and H1c in ESCs, including depletion at active promoters and enrichment at major satellites (Figure 3.7, Figure 3.22, Figure 3.28, Figure 3.29, Figure 3.30, Figure 3.31, and Figure 3.32). Similar to H1d and H1c, $H1^0$ also displayed overall positive correlation with H3K9me3 and inverse correlations with GC% and H3K4me3, although the level of correlation was to a lesser extent (data not shown). Furthermore, $H1^0$ enriched regions were significantly under-represented in gene regions but over-represented in distal intergenic regions with 80.1% of $H1^0$ peaks located in these regions (data not shown). Beside major satellites, $H1^0$ also appeared to be enriched at minor satellites and, to a lesser extent, at LINE L1 elements as determined by ChIP-seq and ChIP-PCR (Figure 3.31 and Figure 3.32), suggesting differential binding preferences of $H1^0$ compared with H1d and H1c.

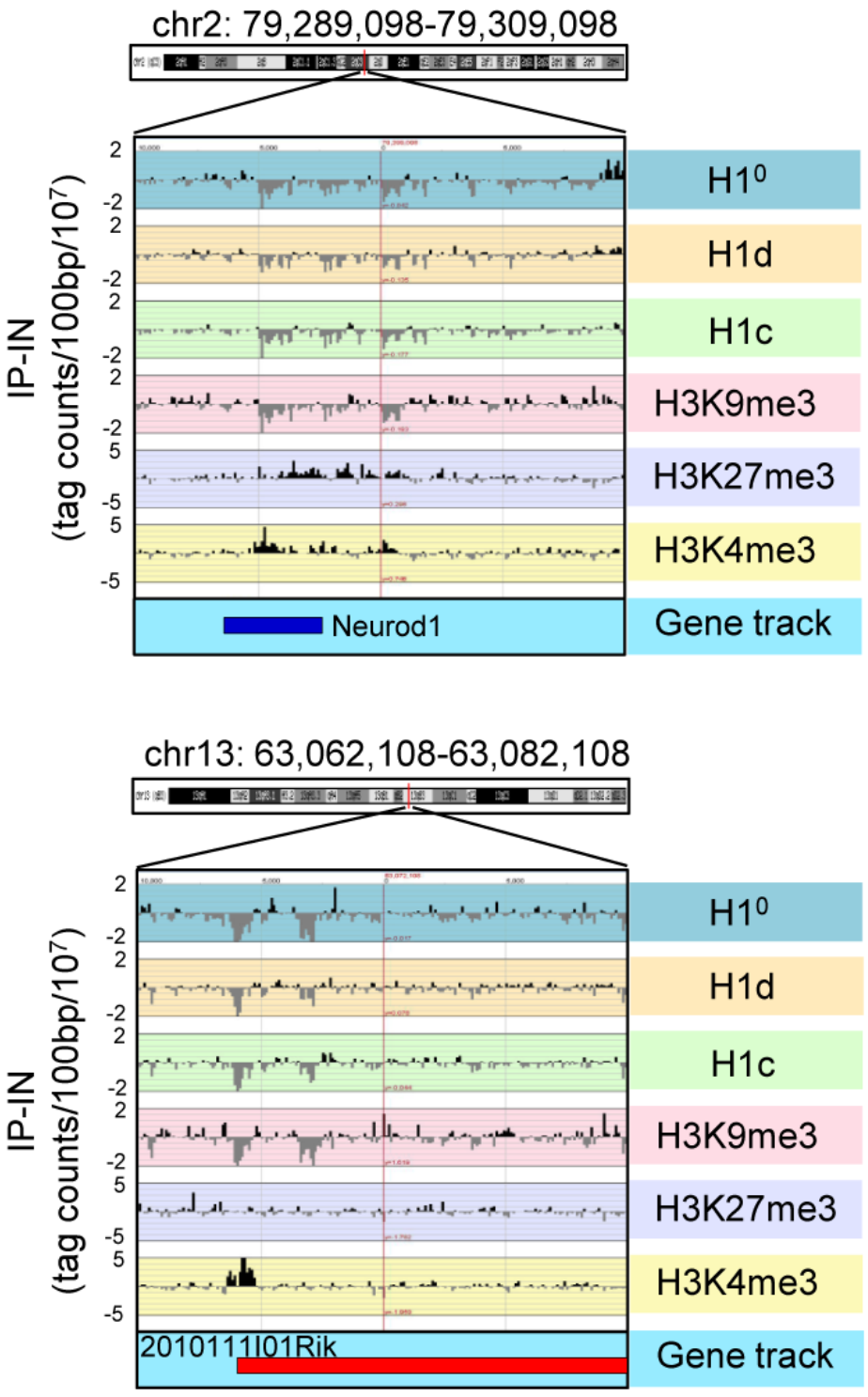

Figure 3.28 Examples of $H1^0$ distribution at TSSs.

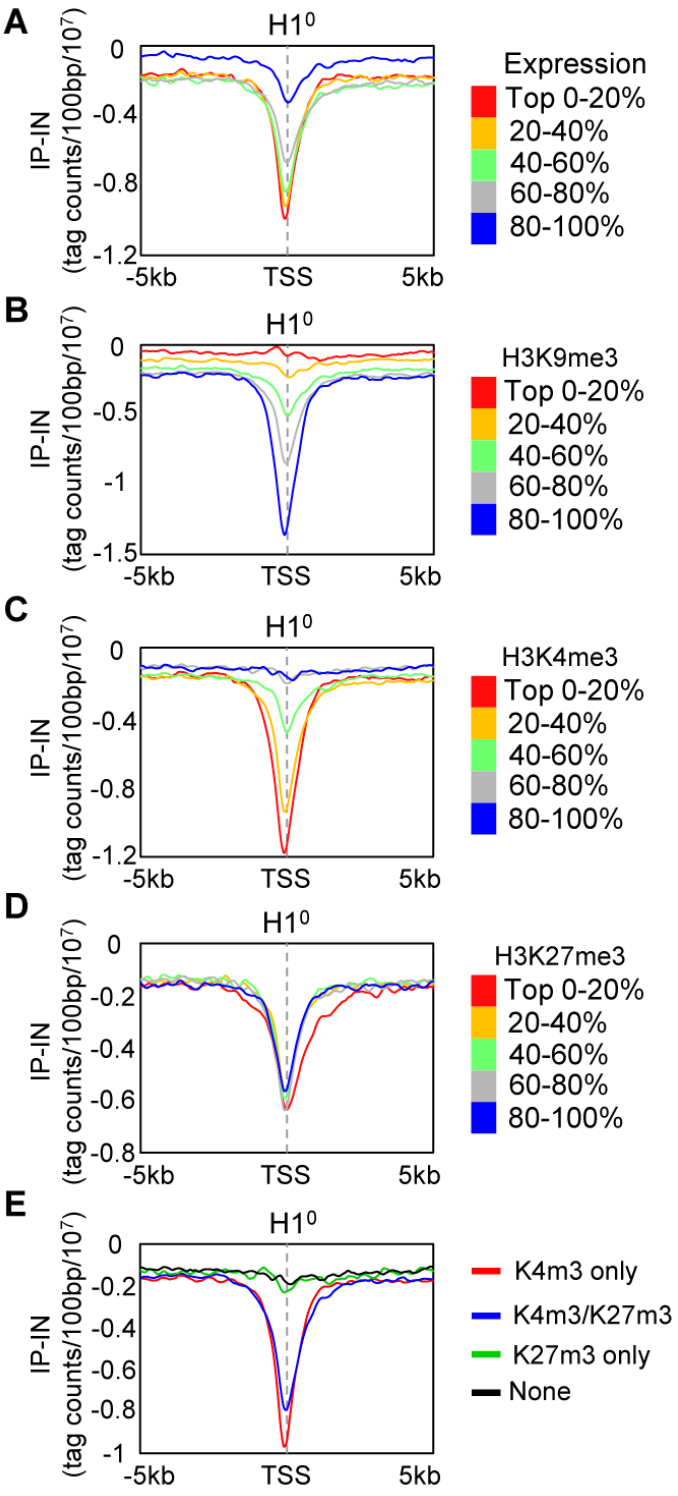

Figure 3.29 $H1^0$ is depleted from active promoters.

Metagene analysis of $H1^0$ levels over a 10 Kb window centered on TSSs partitioned according to the levels of expression (A), H3K9me3 (B), H3K4me3 (C), H3K27me3 (D), and the presence or absence of H3K4me3 and H3K27me3 (E).

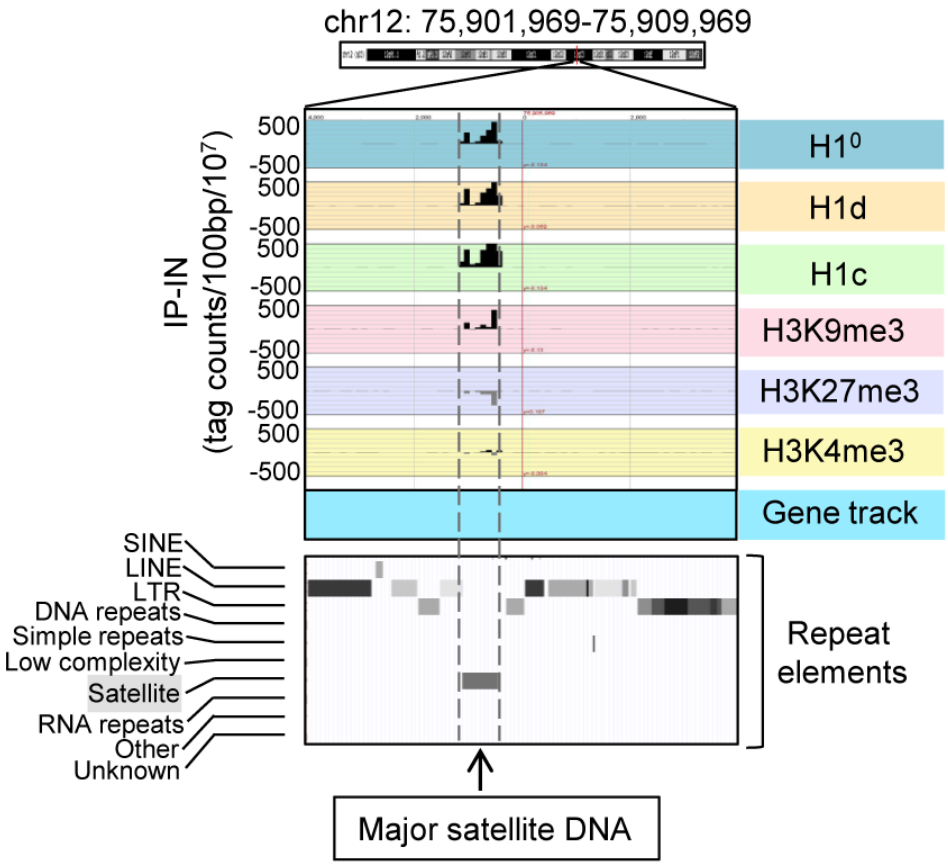

Figure 3.30 A typical peak region of H1^0 at major satellites.

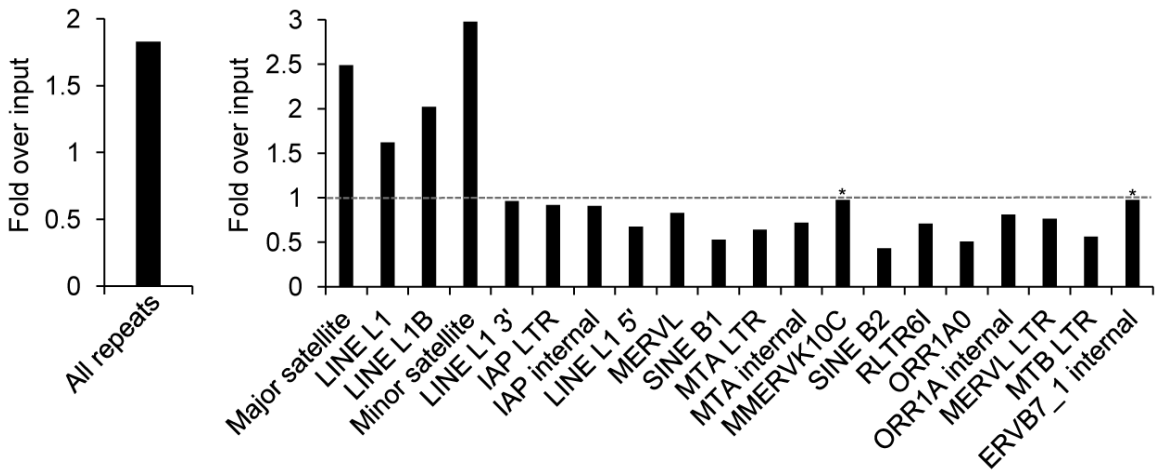

Figure 3.31 H1^{0} is enriched at satellite sequences.

Fold enrichment of percent mappable repeats from the H1^{0} ChIP-seq library over that of input-seq library on total repeats (left) and 20 most abundant repetitive sequences (right). P values calculated with Fisher's exact test comparing ChIP-seq with input-seq libraries are less than 1.8×10^{-21} for all repeats shown except those marked with "*". *: P>0.01.

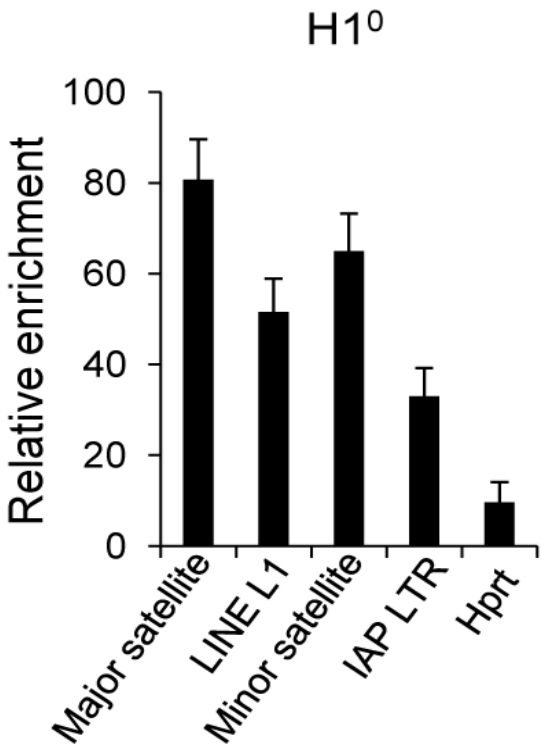

Figure 3.32 qChIP analysis of $H1^0$ occupancy at selected repetitive sequences.
Relative enrichment was calculated by normalizing the signals of ChIP over that of IgG.

3.5 Discussion

H1 Linker histones are abundant chromatin binding proteins that facilitate the formation of higher order chromatin structures (van Holde 1989; Wolffe 1998). The existence of multiple mammalian H1 variants which are differentially regulated during development presumably offers additional levels of modulation on chromatin structure and function. Despite many efforts, the *in vivo* localization and function of individual H1 variants in genome organization remain elusive. Chromatin plays critical roles in stem cell fate determination and reprogramming, and the epigenome of ESCs has been intensively studied. However, the genome-wide maps of one group of the major chromatin proteins, H1 variants, have not been established. Here, we have filled both gaps by generating high resolution maps of three H1 variants in mouse ESCs, identified unique H1 binding features, and discovered an unusual enrichment of H1 variants at major satellites.

On the H1 genome-wide maps we have generated here, H1d and H1c are highly correlated and display similar binding patterns in the ESC genome. Both variants are enriched at AT-rich regions, gene deserts and major satellites, but are depleted at GC-rich, gene-rich regions and especially at active promoters. Thus, despite their differences in compacting DNA *in vitro* and the expression patterns during development (Lennox and Cohen 1983; Wang et al. 1997; Clausell et al. 2009), H1d and H1c are quite similar in overall distribution in the genome, which we surmise contributes to the redundancy among the major somatic H1s as suggested from previous studies of single or double H1 variants knockout mice (Fan et al. 2001; Fan et al. 2003). Nevertheless, analyses of the

regions that are uniquely enriched for H1d or H1c reveal some differences in sequence features (Figure 3.15 and Figure 3.16). H1c has a higher enrichment at major satellites than H1d but is relatively depleted from LINE sequences (Figure 3.15 and Figure 3.19). In addition, H1c enriched regions have a higher proportion in gene bodies and proximal regions compared with H1d peak distribution (Figure 3.11). These differences may account for an additional level of modulation and fine-tuning of genome function by the presence of multiple H1 variants in mammals.

$H1^0$, the H1 variant associated with differentiation, has unique expression pattern and biochemical properties. It is highly basic, expressed in differentiated cell types, and more similar to histone H5 in avian red blood cells than any other somatic variants (Zlatanova and Doenecke 1994; Zhang et al. 2012a). However, overexpressed $H1^0$ (in $fH1^0$ cells) shares the distinctive features of H1d and H1c in ESCs in genome-wide occupancy. It is worth noting, though, that endogenous $H1^0$ proteins are present at very low levels in undifferentiated WT ESCs and the genome-wide localization of $H1^0$ in ESCs may differ significantly from its binding patterns in differentiated cells. It would be interesting to systematically determine the genome-wide maps of histone variants in different cell types, particularly in light of a recent study reporting a distribution pattern change of H1.5 in cellular differentiation (Li et al. 2012). The cell lines and mouse models generated in this study will greatly facilitate these future studies.

The prevalent H1 variants binding with local troughs at active promoters we observed here in the mouse ESC genome is reminiscent of the previous results when ChIP-chip and a pan-H1 antibody were used to map H1 on a portion of the human genome in MCF-7 cells (Krishnakumar et al. 2008) or when DamID method was used to

map H1 in *Drosophila* cells (Braunschweig et al. 2009). The depletion of H1 at TSSs of active genes observed in three systems suggests that this feature is common to all H1s and evolutionarily conserved. However, our study differs from the two previous studies and offers more opportunities for high resolution and in-depth analysis because the knock-in system generated in this study allows for robust and highly specific mapping of H1 variants and deep-sequencing covers the entire genome including the repetitive genome. Furthermore, we have found that the depletion of H1 at active genes is not restricted to regions around the TSS, but also expands to the entire gene encompassing domain (Figure 3.6). Such phenomena suggests that a wide-spread change in higher order chromatin structure may be associated with gene expression and that gene-rich domains may adopt an overall decondensed chromatin structure with less H1 occupancy.

Correlation analyses indicate that H1d and H1c are inversely correlated with GC content, H3K4me3 mark, but positively correlated with H3K9me3 mark across the mouse ESC genome (Figure 3.13). Our finding that the common peaks of H1d and H1c are enriched with AT-rich DNA sequences *in vivo* resonates with the previous observation that H1 is preferentially associated with scaffold associated regions (SAR) (Izaurralde et al. 1989), which are also AT-rich sequences (Mirkovitch et al. 1984). This binding feature may reflect a higher affinity of H1 to AT-tracts observed in *in vitro* studies (Kas et al. 1989; Bonnefoy et al. 1999). The GC content has been suggested to be an intrinsic factor for nucleosome occupancy (Tillo and Hughes 2009), and our data suggest that it may also have an impact on H1 binding. It is also noteworthy that, compared with gene expression levels, H3K4me3 and H3K9me3 correlate better with H1 levels at TSS. For example, we did not observe dips of H1d and H1c around promoters of 40% genes when

partitioned by H3K4me3 or H3K9me3 signals, whereas a small H1 signal dip exists even for the 20% genes with lowest expression values (Figure 3.6, Figure 3.9A, and 3.9B). It is possible that the steady state level of RNA messages (expression) may not faithfully reflect the active/inactive state of the promoters which may correlate better with the status of histone marks. It has been reported that promoters of many genes with low expression have high H3K4me3 levels (Bernstein et al. 2006), and we surmise that H1 may be absent from these gene promoters as well.

The co-localization of H1d and H1c with H3K9me3 suggests that these two variants are enriched at heterochromatin and may facilitate the maintenance of constitutive heterochromatin structure. Such association may be mediated through HP1, the heterochromatin protein binding to H3K9me3 and H3K9 methyltransferase Suv39h and facilitating spreading of heterochromatin marks (Jones et al. 2000; Bannister et al. 2001; Lachner et al. 2001). Indeed, H1 has been shown to interact *in vitro* with HP1α (Nielsen et al. 2001; Daujat et al. 2005). On the other hand, localization of HP1 is impaired in H1 depleted *Drosophila* (Lu et al. 2009b), suggesting that H1 may also contribute to the proper targeting of HP1.

Surprisingly, we found that, at major satellite sequences, H1d and H1c signals are dramatically overrepresented, and this accounts for almost all the increased proportion of H1 sequence reads at repetitive sequences. The levels of H1d and H1c at major satellites are much higher than H3K9me3 (Figure 3.19), a repressive histone mark also enriched at these repeats (Mikkelsen et al. 2007). The enrichment of H1 at major satellites could not be solely attributed to the relatively high affinity of H1c and H1d to AT-rich sequences. Major and minor satellites sequences contain approximately 65% of A and T, with a ratio

of A:T being respective 2.5:1 and 1.8:1. This could result in major satellites having more A-tracts to which H1 might have a higher affinity. Phased nucleosome positioning observed at the major satellites (Zhang and Horz 1984; Linxweller and Horz 1985) could also contribute to the preferential binding of H1 at this region because different nucleosome positioning patterns have been shown to differentially affect H1 binding *in vitro* (Panetta et al. 1998).

In summary, we report high resolution maps of two abundant somatic H1 variants and the replacement H1 variant in mouse ESCs, connecting this important yet under-explored repressive mark with the well-studied ESC epigenome. The enrichment of H1d, H1c and $H1^0$ on major satellites highlights an interesting phenomenon of these H1 variants in the repetitive genome of ESCs, implying possible functions of H1 variants at pericentric chromatin.

CHAPTER 4

H1 LINKER HISTONE REGULATES CHROMATIN STRUCTURE AND FUNCTIONS OF MAJOR SATELLITE REPEATS IN MOUSE ESCS

The results of Chapter 4 have been published in the below article:

Cao K., Lailler N., Zhang Y., Kumar A., Uppal K., Liu Z., Lee E. K., Wu H., Medrzycki M., Pan C., Ho P. Y., Cooper G. P. Jr., Dong X., Bock C., Bouhassira E. E., Fan Y. (2013) High-Resolution Mapping of H1 Linker Histone Variants in Embryonic Stem Cells. PLoS Genet. 9(4): e1003417

4.1 Abstract

H1 linker histone regulates local chromatin structure and the expression levels of specific genes in multiple organisms. Previous results indicate the enrichment of three H1 variants at major satellite repeats in ESCs, while the roles of such enrichment on chromatin structures and functions remain elusive. Here, we show that major satellite repeats display increased nucleosome spacing compared with bulk chromatin in ESCs. Furthermore, whlie deletion of H1c, H1d and H1e has little effect on the levels of histone modifications and DNA methylation at major satellites, H1 depletion causes pericentric chromocenter clustering and de-repression of pericentric repeats. These results identify significant changes at pericentric heterochromatin upon depletion of H1 linker histone, revealing a novel link between H1 and functions of repetitive sequences *in vivo*.

4.2 Introduction

Major satellites are tandem repeats located at mouse pericentric regions with a 234 bp unit repeating up to several megabase pairs (Choo 1997) (Figure 4.1).

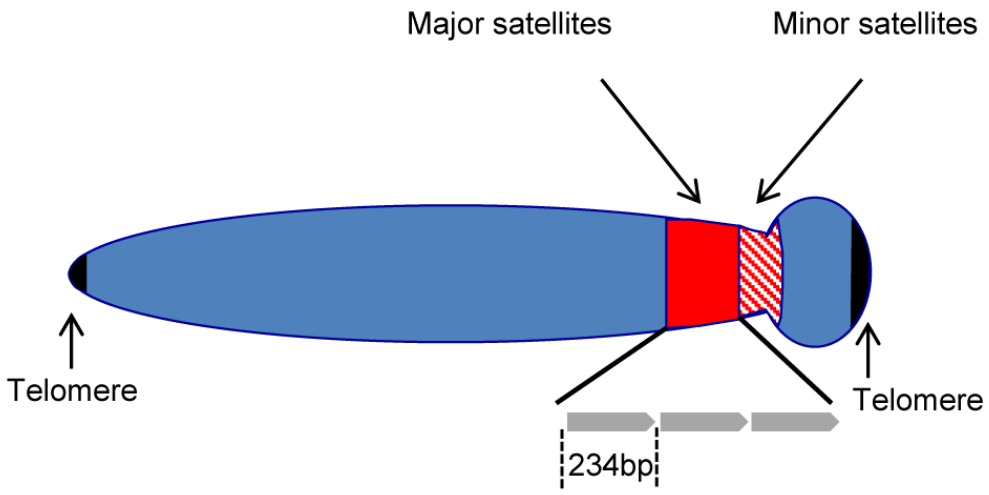

Figure 4.1 Schematic representation of pericentric major satellites, centromeric minor satellites, and telomeres on a mouse chromosome.
The repeating units of major satellites are depicted as grey arrows.

Pericentric major satellites from different chromosomes cluster together to form chromocenters in interphase nuclei (Guenatri et al. 2004), which is an important structural network in cells. Multiple repressive epigentic marks such as H3K9 methylation, H4K20me3, and HP1 are enriched at major satellite repeats (Peters et al. 2001; Guenatri et al. 2004; Martens et al. 2005), which are also heavily methylated (Lewis et al. 1992). However, the pericentric heterochromatin is not transcriptionally silent as noncoding RNAs from major satellites were detected in different organisms (Lehnertz et al. 2003; Probst and Almouzni 2011). These pericentric transcripts seem to

play a role in HP1 recruitment, heterochromatin formation, and mouse development (Maison et al. 2002; Probst et al. 2010; Santenard et al. 2010; Maison et al. 2011). H1 linker histone has been found to interact with HP1 (Nielsen et al. 2001; Daujat et al. 2005; Hale et al. 2006), and our results in Chapter 3 demonstrates a significant enrichment of multiple H1 variants at major satellite repeats, suggesting possible functions of H1 at pericentric heterochromatin.

Several lines of evidence have linked H1 with repetitive sequences. $H1^0$ has been reported to bind at Alu sequences with highest affinity among three H1 variants *in vitro* (Lukiw et al. 1989). H1 is enriched at rDNA repeats in yeast *Saccharomyces Cerevisiae* as shown by ChIP assay (Freidkin and Katcoff 2001). H1 has been shown to regulate genes specifically rather than globally in different systems (Shen and Gorovsky 1996; Steinbach et al. 1997; Hellauer et al. 2001; Fan et al. 2005; Ni et al. 2006; Sancho et al. 2008; Lu et al. 2013). Interestingly, H1 depletion leads to an activation of a group of transposable elements In *Drosophila* (Lu et al. 2013). Multiple mechanisms including DNA methylation and histone modifications are involved in H1 regulated transcription (Fan et al. 2005; Lu et al. 2013; Yang et al. 2013). Nevertheless, the localization and function of H1 on repetitive genome in mammals remain under-explored.

Here, we demonstrate that major satellites display a longer nucleosome repeat length than bulk chromatin in ESCs. By comparing WT and H1 TKO ESCs, we show that H1 depletion leads to chromocenter clustering and an increased expression of major satellites independent of multiple epigenetic marks at these regions.

4.3 Materials and Methods

4.3.1 Determination of nucleosome repeat length (NRL)

ESC nuclei were extracted and MNase digestion was performed as described previously (Fan et al. 2005). Briefly, 2.5×10^6 nuclei were resuspended in 200 μl of MNase digestion buffer (0.32 M sucrose, 50 mM Tris-HCl pH 7.5, 4 mM $MgCl_2$, 1 mM $CaCl_2$, 0.1 mM PMSF) and digested at 37℃ with 20 units of micrococcal nuclease (MNase) (Worthington) for time course analysis or 2 units of MNase (Worthington) for 5 min in analysis shown in Figure 4.4. Nuclei were lysed and DNA was subsequently purified and analyzed by electrophoresis. Southern blotting was performed using major or minor satellite specific probes as described previously (Fan et al. 2005). The NRL at each time point was calculated using the regression line generated with size (bp) of polynucleosomes (Fan et al. 2003; Fan et al. 2005), and the values at time "0" were extrapolated as described previously (Gilbert and Allan 2001).

4.3.2 Fluorescence in situ hybridization (FISH)

FISH was performed as described previously (Zhou et al. 2005). The major satellite probe was biotin-labeled, denatured and hybridized to the slides overnight. The nuclei were incubated with FITC-Avidin for 1 hour, and counterstained with DAPI. Signals were detected with an Olympus Epifluorescence Microscope (Olympus, Inc.) equipped with an Olympus QCLR3 cooled digital camera. The experiments were repeated three times, and the number of chromocenters for each cell line was counted by three researchers as blind tests. Statistical analysis was performed using a Mann-

Whitney U nonparametric test. Areas of chromocenters were quantitated using AxioVision software V4.8.2.0 and presented as pixel^2. The conversion factor of pixel/micron was 18.7 pixels per micrometer.

4.3.3 Quantitative reverse transcription-PCR (qRT-PCR)

1 μg of total RNA extracted from ESCs was treated with RNase free DNaseI (Sigma-Aldrich) and reverse transcribed using a SuperScript first-strand cDNA synthesis kit with random hexamers (Life Technologies). Triplicate PCR reactions using the iQ SYBR Green Supermix (Bio-Rad) were analyzed in a MyIQ Real-Time PCR Detection System (Bio-Rad). All samples were typically analyzed in two independent experiments. Relative expression units were calculated by subtracting the mock reverse-transcribed signals (RT-) from reverse transcribed signals (RT+) and normalizing the adjusted values with signals of the housekeeping gene *GAPDH*. The qRT-PCR primers for repetitive sequences are the same as in qChIP, and the primers for *GAPDH* are as described previously (Zhang et al. 2012a) and listed in Table A.1.

4.3.4 Bisulfite treatment of DNA and sequencing analysis

1 μg of DNA extracted from ESCs was treated with the CpGenome DNA modification Kit (Millipore) according to the manufacturer's manual. 20 ng of treated DNA was used in each PCR reaction as previously described (Fan et al. 2005). The primers used to generate PCR products from the bisulfite-converted DNA are specific for the converted DNA sequence of the analyzed regions. The PCR products were subsequently cloned using the TOPO TA Cloning kit (Life Technologies), and colonies

containing the converted DNA inserts were picked. DNA inserts were sequenced and analyzed with BiQ Analyzer (Bock et al. 2005). Primers for major and minor satellites were as previously described ((Yamagata et al. 2007) and Table A.1).

4.4 Results

4.4.1 Increased nucleosome repeat length at major satellite sequences in ESCs

The level of H1 has been shown to be a determinant of nucleosome repeat length (NRL) with a higher level of H1 correlating with a longer NRL (Woodcock et al. 2006; Oberg et al. 2012). To validate the enrichment of H1 variants at major satellites and to investigate its impact on the local chromatin structure at these regions, we measured the NRL of bulk chromatin and that of the pericentromeric (major satellites) and centromeric (minor satellites) regions with a time-course micrococcal nuclease (MNase) digestion assay. Southern blotting images revealed that chromatin at major satellites was more resistant to MNase digestion than bulk chromatin and minor satellites (Figure 4.2). Consistent with previous studies (Fan et al. 2005), the bulk chromatin of mouse ESCs displayed a NRL of ~187 bp (Figure 4.3). However, the NRL at major satellites had a value of 200 bp, which was ~13 bp and ~8 bp longer than the NRLs of respective bulk chromatin and minor satellites in ESCs (Figure 4.3). These results suggest that the enrichment of H1d and H1c at major satellite repeats may contribute to the increase of NRL in the pericentromeric region compared with bulk ESC chromatin. Analysis of H1c/H1d/H1e triple knockout (H1 TKO) ESCs established previously, which have an H1/nuc ratio of 0.25 in bulk chromatin compared with that of 0.46 in WT ESCs (Fan et al. 2005), indicated that H1 depletion caused a proportional decrease of NRLs in bulk chromatin, major satellites and minor satellites (Figure 4.4 and Figure 4.5). Consistently, qChIP analysis using a pan-H1 antibody showed total H1 levels were reduced at major and minor satellites by H1 depletion (Figure 4.6).

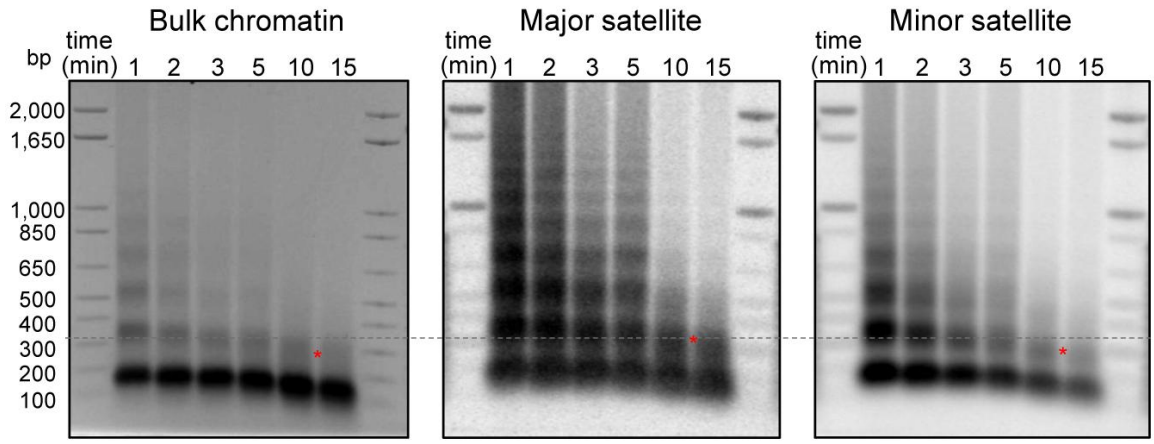

Figure 4.2 Nucleosome repeat length analyses in WT ESCs
DNA isolated from ESC nuclei digested with MNase at different time points were analyzed by ethidium bromide (EB) -stained gel (left), transferred to membrane which was sequentially probed with major satellites (middle) and minor satellites (right) using Southern blotting. The positions of di-nucleosomes with 10-minute MNase digestion are marked by *. The dashed line indicates di-nucleosome position of major satellites, which is higher than that of bulk chromatin and minor satellites.

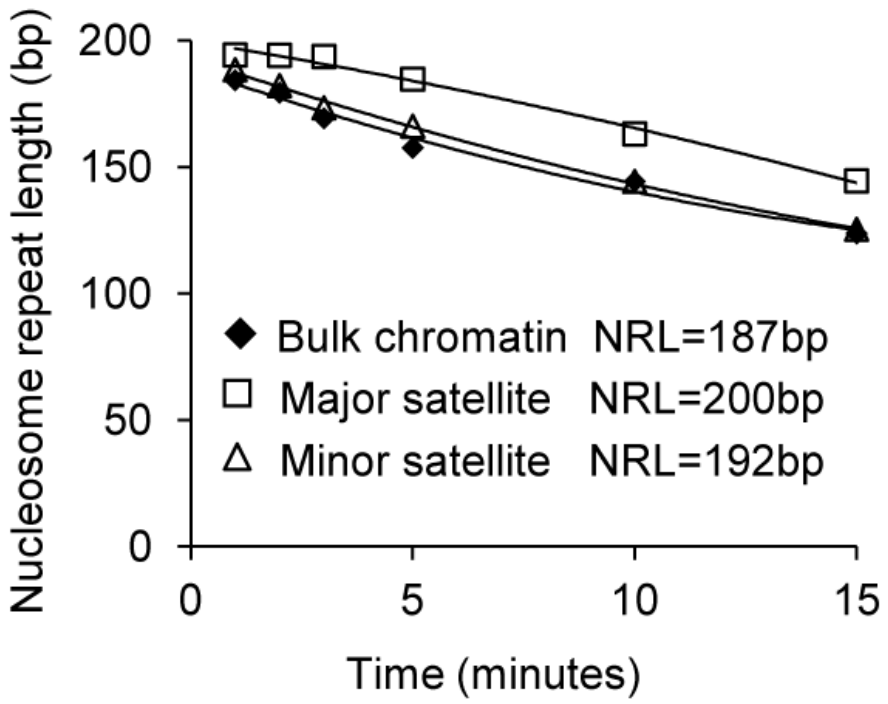

Figure 4.3 Increased nucleosome repeat length at major satellite repeats in ESCs.
The NRLs were calculated from the images presented in Figure 4.2 by extrapolating the corresponding curves to time "0" as described (Gilbert and Allan 2001).

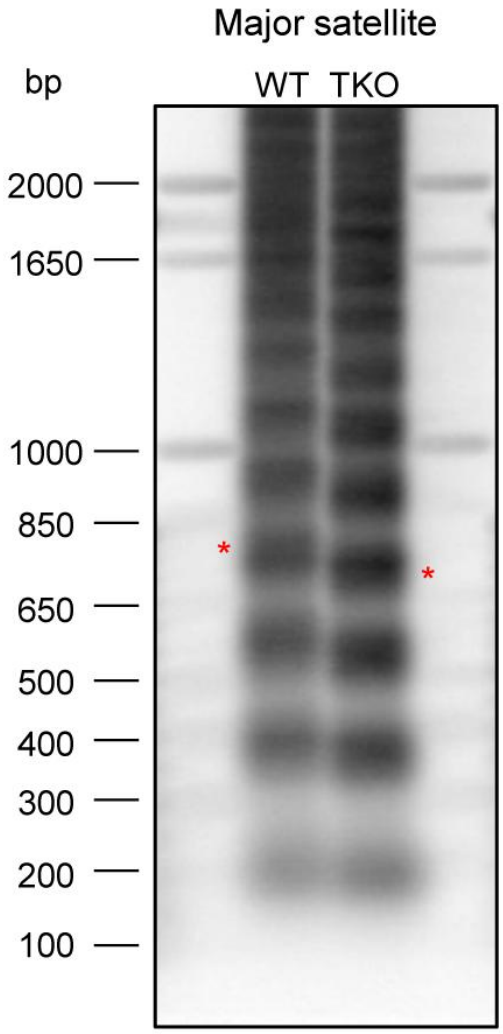

Figure 4.4 Southern blotting analysis of partially digested nuclei using a major satellite probe.
The tetra-nucleosome bands are indicated by asterisks.

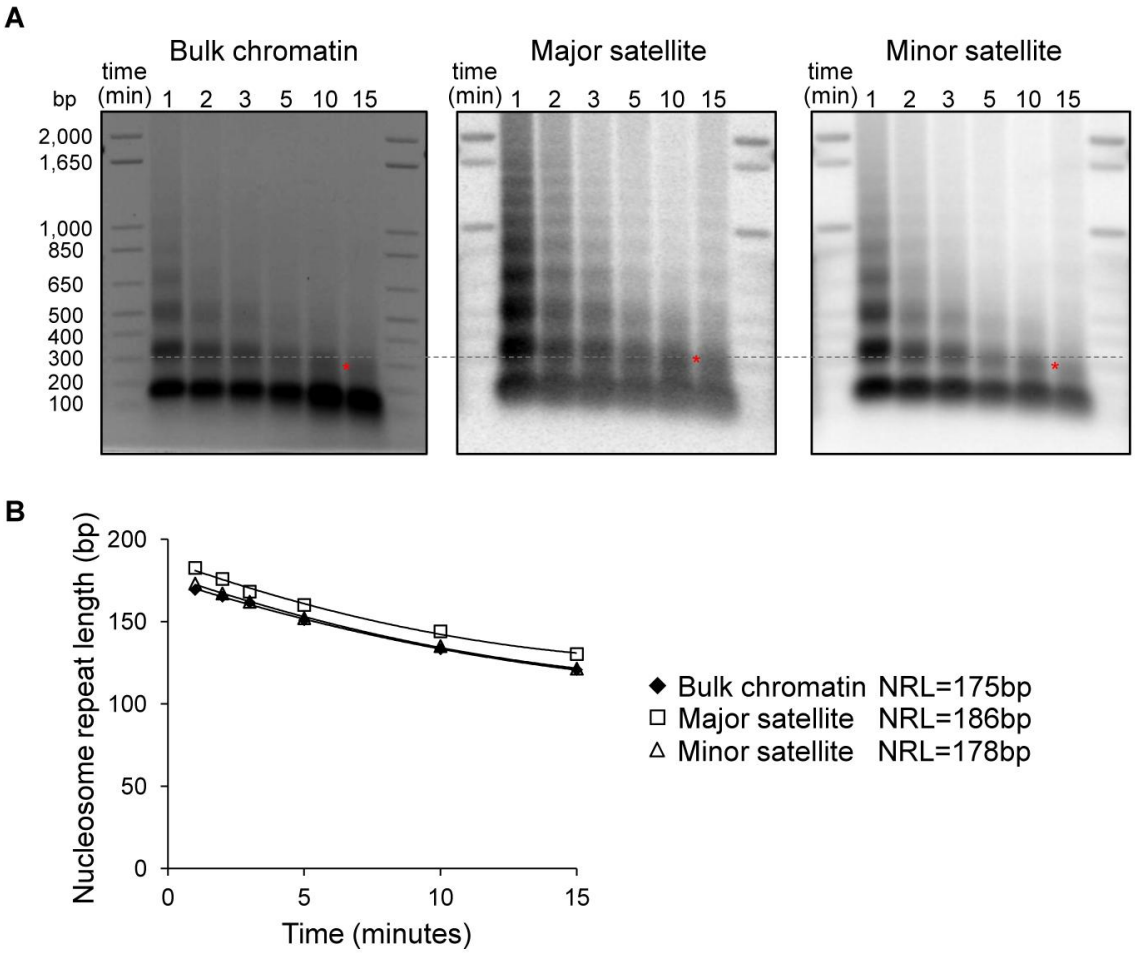

Figure 4.5 Elevated NRLs at major satellites compared with bulk chromatin and minor satellites in H1 TKO ESCs.

Data from EB-stained gel image (A, left) and corresponding Southern blots (A, middle and right) are plotted in (B). The positions of di-nucleosome with a 10-minute MNase digestion are marked by asterisks in (A). The dashed line in (A) indicates the di-nuleosome position in major satellites, which is higher than that of bulk chromatin and minor satellites. NRLs in (B) were calculated by extrapolating the corresponding curves to time 0 as described (Gilbert and Allan 2001).

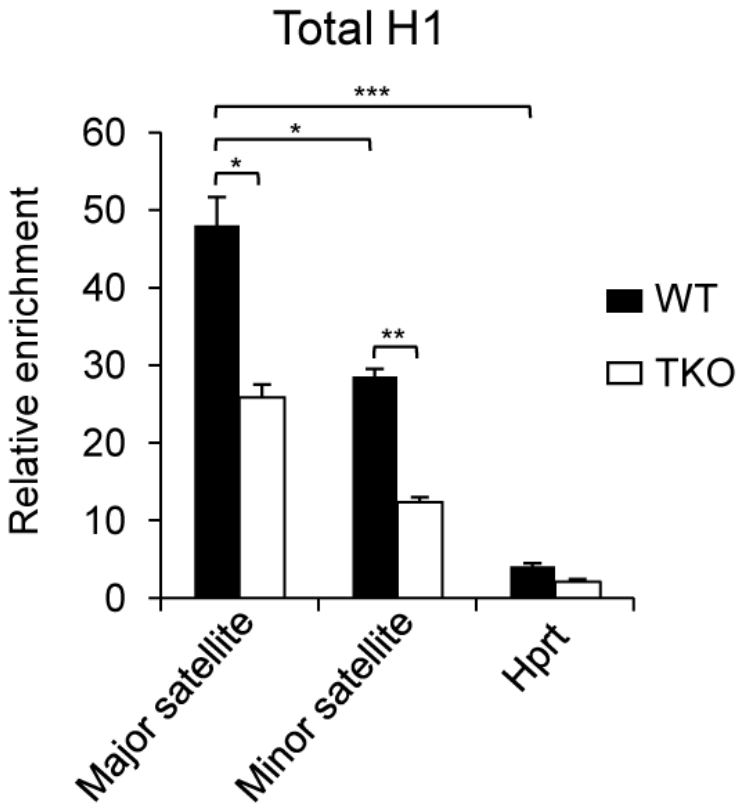

Figure 4.6. H1 depletion leads to reduced H1 occupancy at major and minor satellites.
qChIP analysis of H1 occupancy at major satellites, minor satellites, and *HPRT* gene in WT and H1 TKO ESCs. ChIP signals over IgG levels are presented as mean ± S.D. *: P<0.05; **: P<0.01; ***: P<0.001.

4.4.2 H1 depletion leads to chromocenter clustering

Major satellite repeats at pericentric heterochromatin from different chromosomes tend to cluster together and form the chromocenter, a nuclear compartment that plays an important role in structural maintenance of the chromosomes (Hsu et al. 1971; Guenatri et al. 2004). Several chromatin proteins such as MeCP2, MBD2, DNMT3a, DNMT3b, and UHRF1 have been shown to contribute to chromocenter clustering (Brero et al. 2005; Gilbert et al. 2007; Papait et al. 2008), however, the role of H1 in chromocenter formation has not been studied to date. Since both H1d and H1c are markedly enriched at major satellites, we set out to determine the effects of H1 depletion on chromocenter clustering in WT and H1 TKO ESCs. To this end, I collaborated with Yunzhe Zhang and Dr. Zheng Liu for fluorescence in situ hybridization (FISH) analysis using a major satellite specific probe in these cells. The chromocenter numbers in H1 TKO ESCs (median=8, n=160) were significantly lower than WT cells (median=17, n=206) (Figure 4.7), and the size of chromocenters in H1 TKO ESCs on average was bigger than that in WT ESCs (Figure 4.8), demonstrating a previously unnoticed defect in the pericentromeric chromatin structure caused by H1 depletion. Analysis of "rescue" (RES) cells established previously (Zhang et al. 2012a) showed that overexpressing H1d in H1 TKO cells effectively restored the size and the numbers of chromocenters to the levels comparable to WT cells (Figure 4.7 and Figure 4.8). Similarly, $H1d^{FLAG}$ and $H1c^{Myc}$ cells displayed normal chromocenter clustering as WT ESCs (Figure 4.9). We note that the level of $H1^0$, the replacement H1 variant, was increased significantly in TKO ESCs compared with that in undifferentiated WT ESCs where $H1^0$ was minimal (Fan et al. 2005; Zhang et al. 2012a). To examine if the increased chromocenter clustering in H1

TKO ESCs could be attributed to an increase in $H1^0$ levels, we performed DNA FISH in $fH1^0$ ESCs. Such analysis indicated that the chromocenter numbers were not reduced compared with WT ESCs (Figure 4.9). These results indicate that the increased chromocenter clustering is likely due to the dramatic decrease of total H1 levels in H1 TKO ESCs.

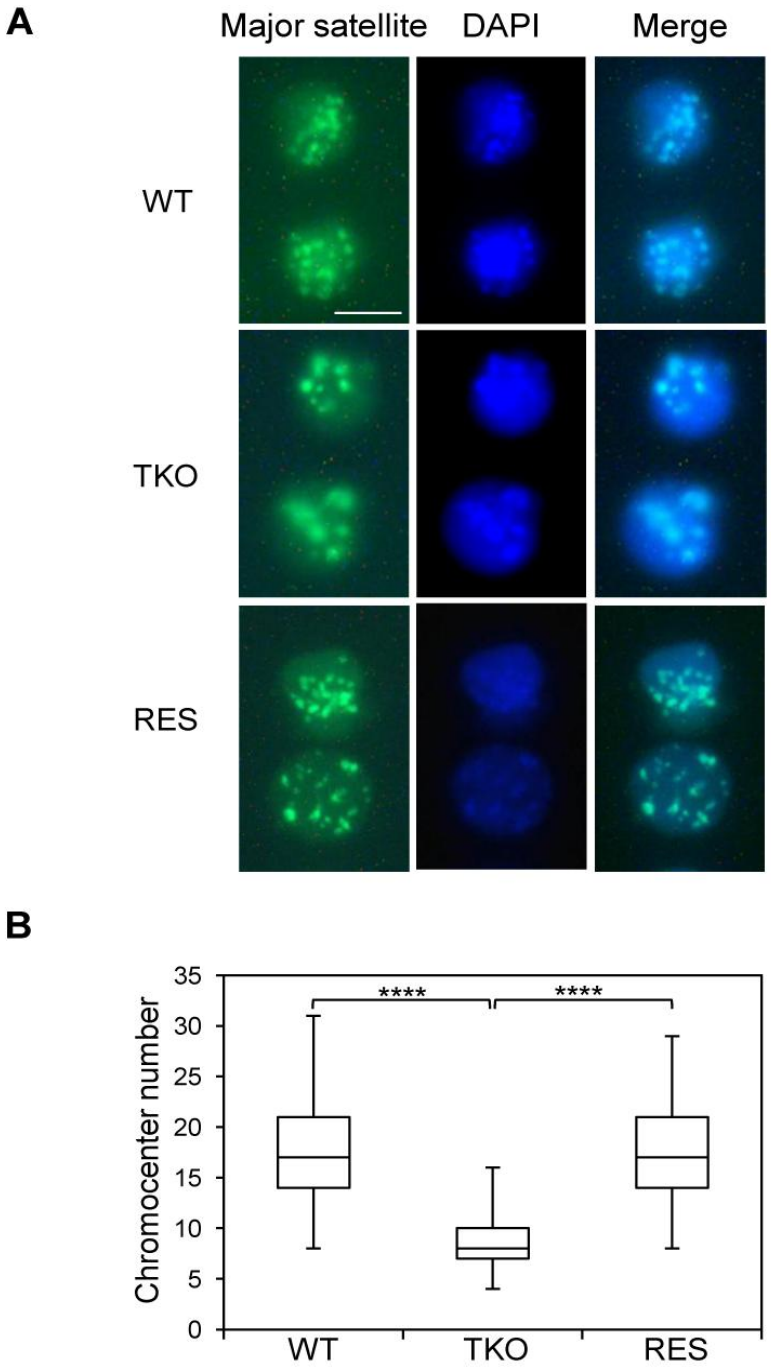

Figure 4.7 H1 depletion leads to chromocenter clustering.
(A) Typical images of WT (top), H1 TKO (middle), and RES ESCs (bottom) of FISH with a major satellite probe (left), DNA stain DAPI (middle), and merged images (right). Scale bar: 10 μm.
(B) Box plots of chromocenter numbers in the nuclei of WT, H1 TKO, and RES ESCs. The line in the box indicates the median, while the bottom and top of the boxes are the 25th and 75th percentiles, respectively. ****: P<0.000001.

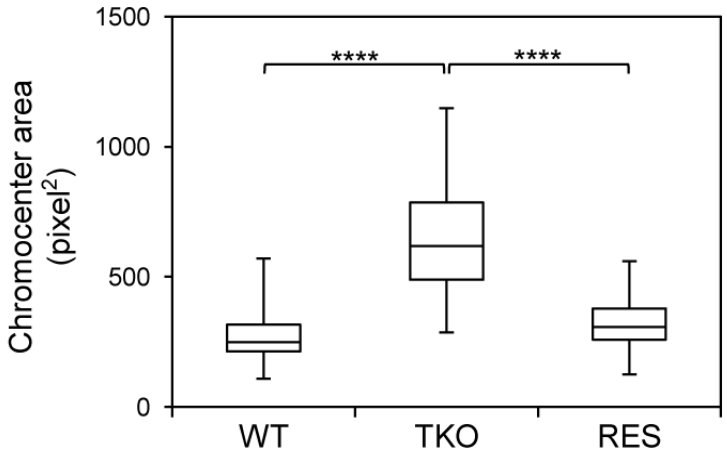

Figure 4.8 Chromocenter area in WT, H1 TKO and RES ESCs.
Chromocenters of 80 nuclei from each cell line were analyzed. The line in the box indicates the median, while the bottom and top of the boxes are the 25th and 75th percentiles, respectively. ****: P<0.000001.

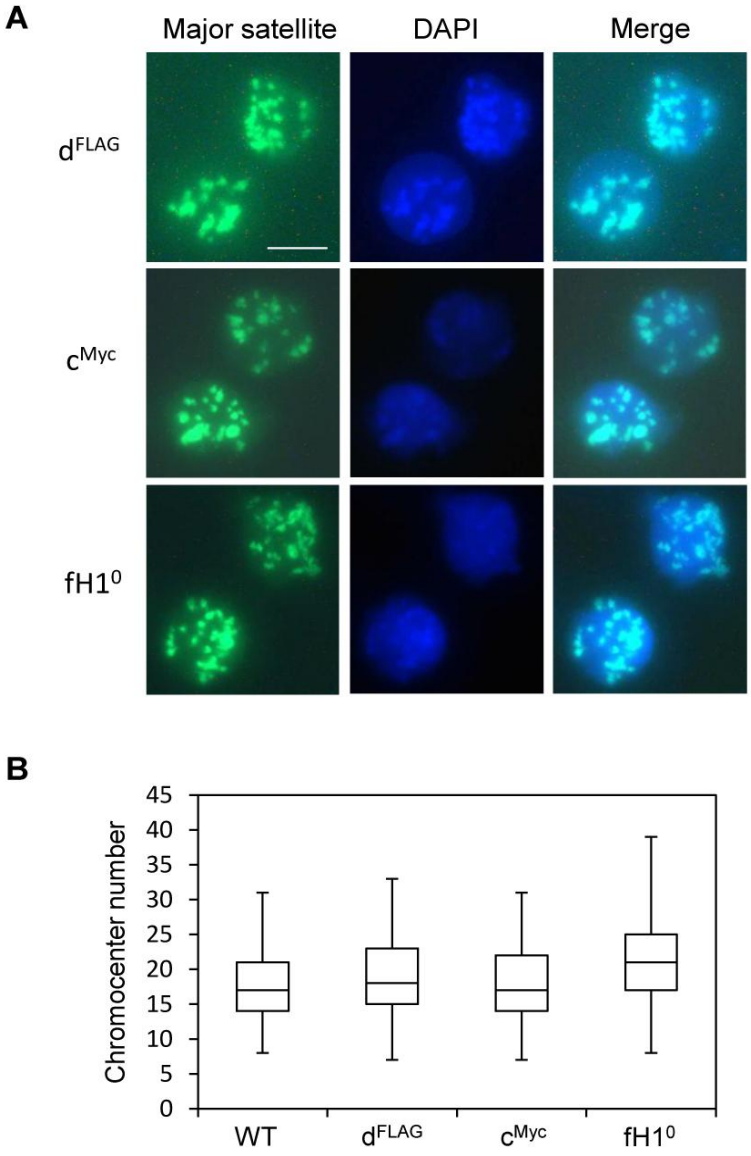

Figure 4.9 FISH analyses of chromocenters in H1d^{FLAG}, H1c^{Myc}, and fH1^{0} ESCs.
(A) Typical FISH images of indicated cells hybridized with a major satellite probe are shown in left panels. DNA was counterstained with DAPI (middle), and merged images are shown in right panels. Scale bar: 10 μm.
(B) Box plots of the numbers of chromocenters in indicated ESCs. The line in the box indicates the median, while the bottom and top of the boxes are the 25th and 75th percentiles, respectively.

4.4.3 H1 regulates the expression levels of major satellite repeats independent of multiple epigenetic marks

Pervasive transcription of repetitive sequences contributes to genome regulation, and aberrant regulation of the expression of satellite sequences interferes with heterochromatin assembly and chromosome segregation (Rudert et al. 1995; Guenatri et al. 2004; Bouzinba-Segard et al. 2006; Berretta and Morillon 2009). To further examine the effects of H1 depletion on major satellites, we analyzed several repetitive sequences for expression and epigenetic marks in WT and H1 TKO ESCs. Quantitative reverse transcription-PCR (qRT-PCR) analysis showed that the expression levels of major satellites were 3.5-fold higher in H1 TKO ESCs than in WT ESCs, whereas the expression levels of minor satellites and LINE L1 were not significantly changed (Figure 4.10). Such de-repression of major satellites by H1 depletion was dramatically curbed in RES cells (Figure 4.10) as well as in $H1d^{FLAG}$ and $H1c^{Myc}$ ESCs (Figure 4.11), indicating that the levels of H1s have a direct impact on transcriptional regulation of major satellites. Notably, the levels of multiple epigenetic marks, such as repressive marks H3K9me3, H3K27me3, and H4K20me3, the active mark H3K4me3, as well as DNA methylation all remained unchanged at the analyzed repeats in H1 TKO ESCs compared with WT ESCs (Figure 4.12 and Figure 4.13). The lack of significant changes in the histone marks and DNA methylation at these repetitive sequences suggests that the increase in expression levels at major satellites may be due to an effect of local chromatin decondensation caused by H1 depletion in H1 TKO ESCs. qRT-PCR analysis indicated that the expression of major satellites remained at low levels in $fH1^0$ ESCs (Figure 4.11), excluding the possibility of $H1^0$ upregulation being responsible for upregulation of major

satellite transcription in H1 TKO ESCs.

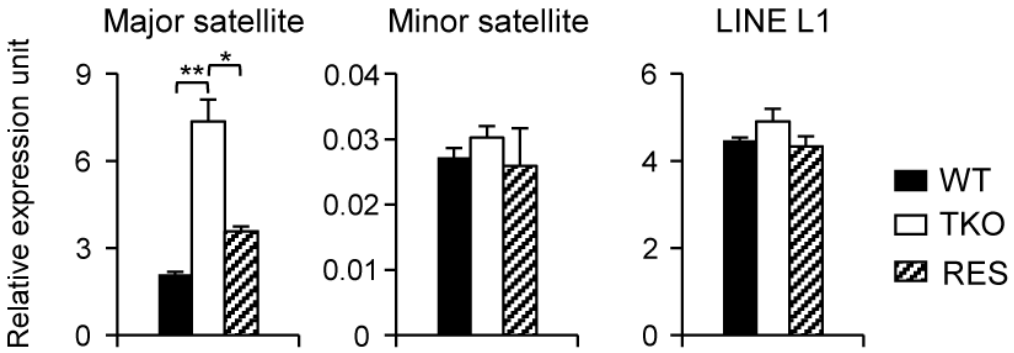

Figure 4.10. H1 depletion leads to increased expression of major satellite repeats. Analyses of expression of selected repeats in WT, H1 TKO, and RES ESCs by qRT-PCR. Data are represented as mean +/- S. D.. *: P<0.05; **: P<0.01.

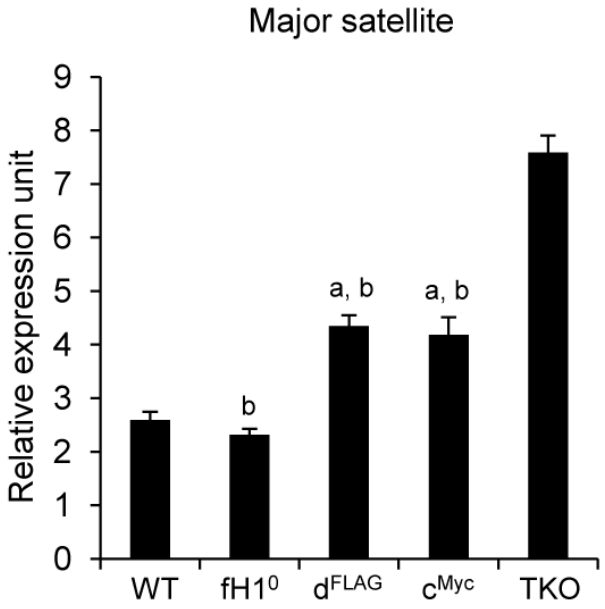

Figure 4.11 Expression analysis of major satellites in fH1^{0}, H1d^{FLAG}, and H1c^{Myc} ESCs.

Data are shown as mean ±S.D. a: $P<0.05$ in comparison with WT; b: $P<0.05$ in comparison with TKO.

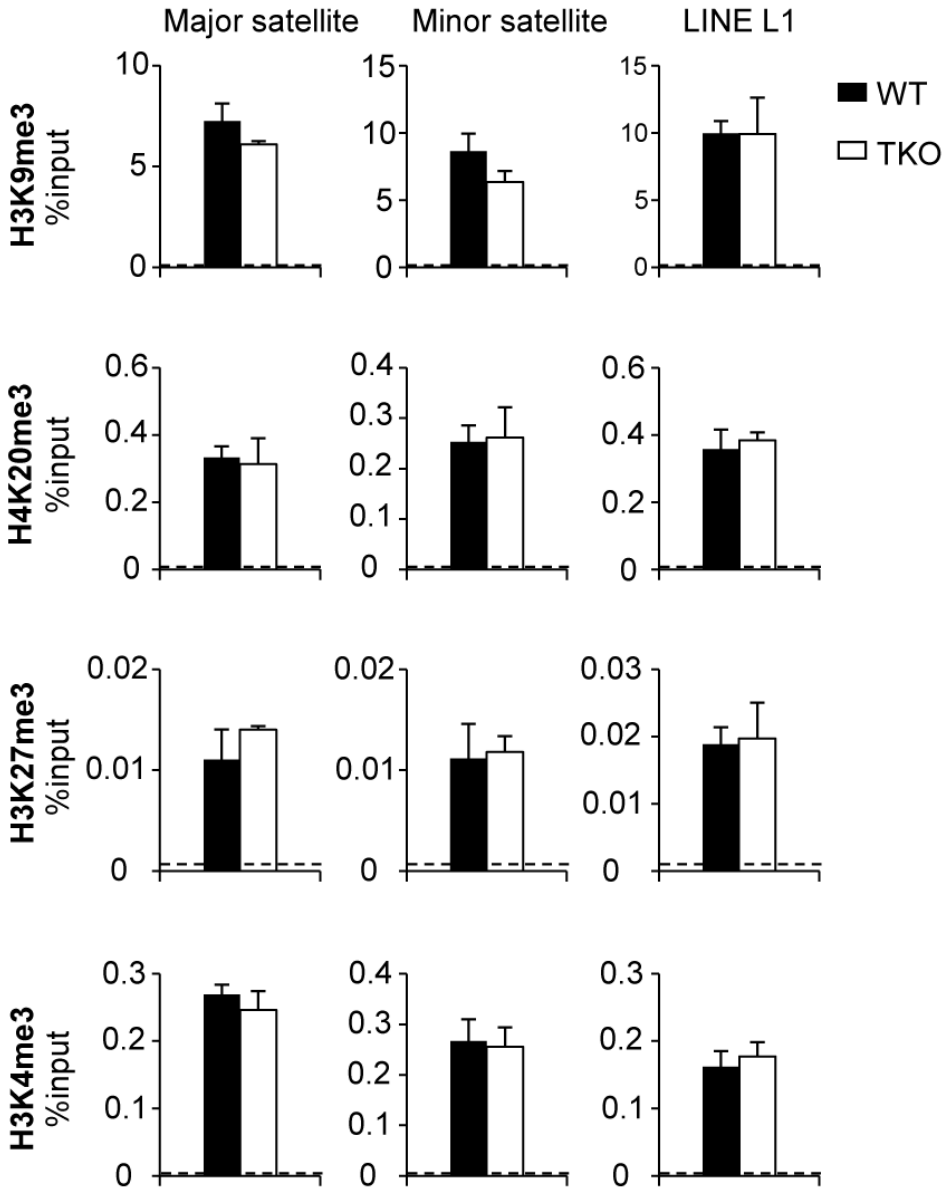

Figure 4.12. Multiple histone marks at major satellites are not affected by H1 depletion.
qChIP analysis of three repressive histone marks and one active histone mark at selected repetitive sequences in WT and H1 TKO ESCs. Dashed lines indicate the highest level of signals detected by qChIP with IgG antibody.

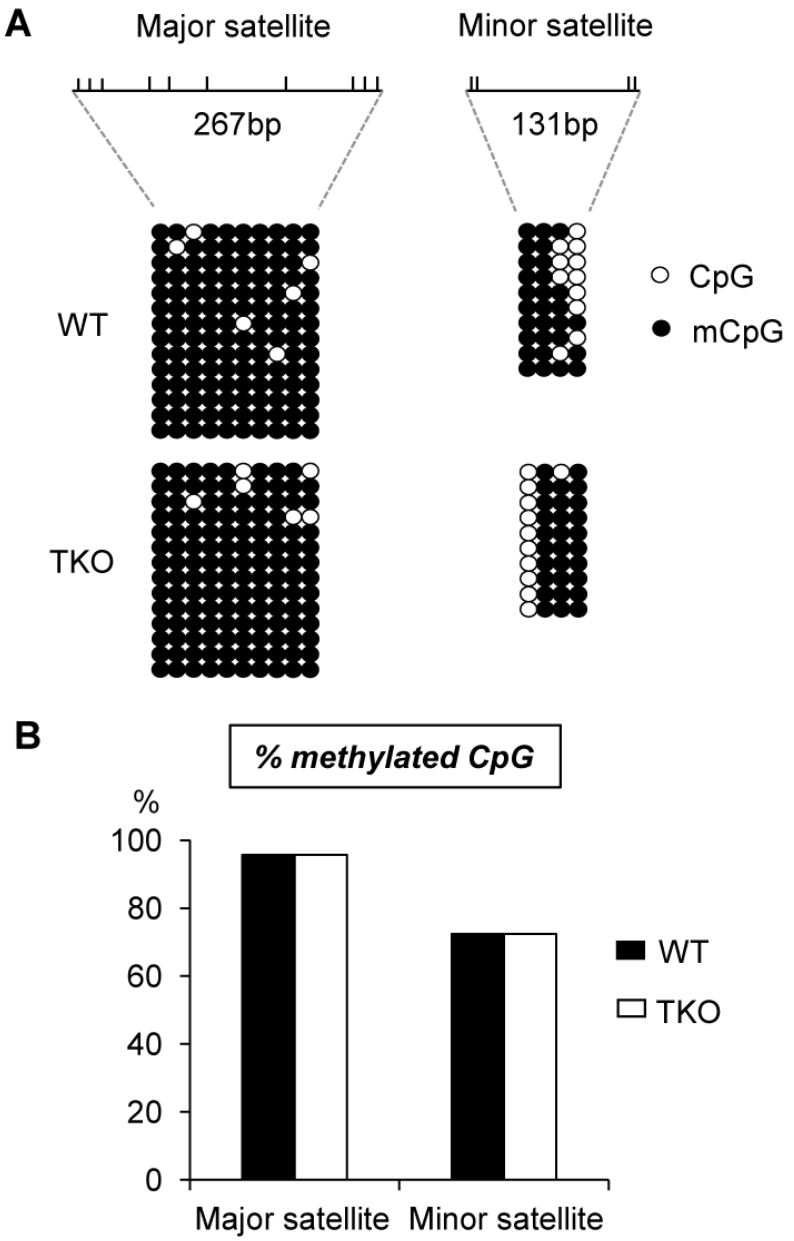

Figure 4.13 DNA methylation at major satellites are not affected by H1 depletion. Bisulfite sequencing analysis (A) and percent of methylated CpG (B) of major, minor satellite sequences. The positions of CpG sites analyzed are marked as vertical ticks on the line.

4.5 Discussion

The overrepresentation of H1 at major satellites in ESCs is supported by a longer NRL (Figure 4.2), which suggests a higher local H1 level than bulk chromatin and minor satellites. Consistent with previous observations (Guenatri et al. 2004; Gilbert et al. 2007), we find that major satellites are more resistant to MNase digestion than bulk chromatin and minor satellites in ESCs (Figure 4.2), suggesting that pericentromeric regions may adopt special higher order chromatin structure as indicated by sucrose sedimentation assay (Gilbert and Allan 2001). High resolution mapping in this study identifies major satellites as the dominant preferential binding sites for H1 variants in ESCs, suggesting that H1 may play an important role in mediating the formation of distinct chromatin structure at pericetromeric regions. This is further supported by the effects of H1 depletion on chromocenter clustering and expression of major satellites. We note that a higher NRL in major satellites than bulk chromatin is also present in H1 TKO ESCs (Figure 4.5), suggesting a possible enrichment of the remaining H1 variants at major satellite sequences in H1 TKO ESCs. Consistently, we have found that overexpressed $H1^0$ also appear to preferentially accumulate at satellite sequences in ESCs.

Mouse major satellites, constituting the pericentromere (Choo 1997; Probst and Almouzni 2008) necessary for chromosome structure and function, are shown to form clusters/chromocenters, exhibit distinct heterochromatin features and adopt a more stable and condensed chromatin conformation than the bulk chromatin (Gilbert and Allan 2001; Guenatri et al. 2004). Our findings of the preferential binding of H1 at major satellites and chromocenter clustering (reduced number of chromocenters) in H1 TKO ESCs

suggest that H1 contributes to and may be required for the proper formation of pericentric heterochromatin. The rescue of the clustering effects by overexpressing H1d in H1 TKO ESCs or in $H1d^{FLAG}$ and $H1c^{Myc}$ cells compared with H1 TKO ESCs indicates that the total H1 level, rather than a specific H1 variant, is a key determining factor of chromocenter clustering. This conclusion is further supported by our finding that overexpressing $H1^0$ level to 3.5 fold of that of endogenous $H1^0$ in WT ESCs has little effect on chromocenter numbers or major satellite expression. *In vitro* studies have shown highly cooperative binding of H1 globular domain to DNA (Thomas et al. 1992), a property which we speculate could contribute to increased chromocenter clustering in the face of marked reduction of H1 levels in H1 TKO ESCs. A larger nucleosome spacing (200 bp) (Figure 4.3) together with a higher local H1 level at major satellites could be important for efficient compaction of pericentromeric chromatin because nucleosome arrays with a NRL of 197 bp are able to form 30 nm fiber structure *in vitro* in the presence of linker histones whereas arrays with a short NRL are only able to form thinner and less compact structures (Routh et al. 2008).

The effects of H1 on major satellites are not restricted to chromatin structure and heterochromatin formation. Loss of H1c, H1d and H1e causes a dramatic increase in transcripts from major satellites, but does not change the levels of the repressive epigenetic marks, H3K9me3, H4K20me3, H3K27me3, or DNA methylation at these sequences. This suggests that the increase in expression of major satellites in H1 TKO ESCs is not mediated by loss of these repressive epigenetic marks, but rather caused by reduced binding of H1 per se or the potential decondensation of local chromatin structure. The phenomenon of changes in chromocenter organization independent of H3K9me3 is

reminiscent of results from deletion of UHRF1 (Papait et al. 2008), a histone binding protein or overexpression of MeCP2 in mouse myoblasts (Brero et al. 2005). Chromocenter organization is likely to be independent of H3K9me3 pathway because double deletion of Suv39h1 and Suv39h2 has minimal effects on the number and size of chromocenters in mouse cells (Peters et al. 2001; Harnicarova Horakova et al. 2010). The expression changes in major satellites in H1 TKO ESCs are also not due to potential changes in cell cycle since H1 TKO ESCs have similar growth rate (Fan et al. 2005) and cell cycle profiles (data not shown) to WT ESCs. The reduction in expression levels of major satellites detected in RES cells compared with H1 TKO cells further supports that the drastic decrease in H1 levels causes de-repression of major satellites. Noncoding major satellite transcripts have been shown to be important for proper chromocenter formation (Probst et al. 2010; Casanova et al. 2013), thus we speculate that the increased levels of major satellite transcripts contribute to chromocenter clustering in H1 TKO cells. In light of previous findings that ESCs null for DNA methyltransferases displayed chromocenter clustering (Gilbert et al. 2007), similar to what we observed in H1 TKO ESCs, we surmise that H1 and DNA methylation may act cooperatively in the proper maintenance of chromocenter structure.

Collectively, the results in this study have demonstrated the increased chromocenter clustering and major satellite transcription by H1 depletion, and suggest important roles of the dominant H1 variants in ESCs in maintaining pericentric chromatin properties.

CHAPTER 5

CONCLUSIONS AND FUTURE STUDIES

This dissertation is focused on revealing the genome-wide binding landscape of three H1 histone variants: H1d, H1c, and $H1^0$, in mouse embryonic stem cells. It is challenging to map specific H1 variants because antibodies of high quality and specificity to individual H1 variants for ChIP assay are lacking. To circumvent the technical difficulty, we have established a knock-in experimental system in mouse embryonic stem cells in which endogenous H1d and H1c are replaced by their corresponding epitope-tagged H1 variant, respectively. Besides ESCs, mice carrying FLAG-tagged H1d at the endogenous H1d locus have been derived. Since $H1^0$ is barely detectable in ESCs, tagged $H1^0$ has been overexpressed to a physiological level in WT ESCs. We have demonstrated that tagged H1s are functionally interchangeable to their endogenous counterparts *in vivo*, thus proving the feasibility of the system. The establishment of these cell lines and the mouse model paved the way for dissecting the role of individual H1 variants in pluripotent stem cells, differentiated cells, as well as in adult tissues.

Compared with other genetic methods for investigating gene functions in mammals, such as overexpressing exogenous proteins with plasmids or lentiviruses, the knock-in approach guarantees that the tagged (or modified) gene of interest is expressed at the endogenous level, circumventing the possible artifacts caused by the abnormal expression. Recent development in genome editing technologies, such as zinc-finger nucleases (ZNFs) (Miller et al. 2007), transcription activator-like effector nucleases (TALENs) (Hockemeyer et al. 2011), and CRISPR-Cas nuclease systems (Wang et al.

2013), could further enhances the efficiency of genome editing. Therefore, such an approach could be used to study other H1 variants or other proteins *in vivo*.

Mapping of H1d and H1c with ChIP-seq indicate that these variants are depleted at active promoters, GC- and gene-rich regions. H1d and H1c are inversely correlated with H3K4me3 and GC content, while positively correlated with H3K9me3 as shown by genome-wide correlation studies. CEAS and EpiGRAPH analyses have identified similarity and differences of H1d and H1c enriched regions. $H1^0$ has overall similar binding patterns with somatic H1 variants. These results characterize the features of the genome-wide H1 localization.

Our results show different binding patterns of H1d, H1c and $H1^0$ in specific genome regions (Figure 3.11), which could provide insights in understanding the modulation of H1 variants in chromatin structure and gene expression. It would be interesting to also map other H1 variants in mESCs, such as H1a, which has the lowest chromatin binding affinity among major somatic H1 variants by FRAP analysis and biochemical assays (Th'ng et al. 2005; Clausell et al. 2009) and a different genome binding pattern compared with other H1 variants in human lung fibroblasts (Izzo et al. 2013).

Future study would be carried out to map the genome-wide binding features of H1 variants at differentiated cell types and tissues since somatic cells have a different chromatin state compared with pluripotent cells (Meshorer et al. 2006). Moreover, the genome-wide correlation of H1 variants and specific histone marks suggest that H1 may regulate the distribution of these marks. Previous studies have shown that H1 depletion leads to aberrant alterations of histone marks at pluripotency genes during ESC

differentiation (Zhang et al. 2012a). ChIP-seq could be utilized to analyze genome-wide distribution of histone marks in differentiating WT and H1 depleted cells. Comparing the profiles of histone modification in these cells will facilitate the deduction of the temporal sequences and loci regulated by histone H1 during the differentiating processes.

Recent discoveries and progresses in induction of pluripotency from somatic cells by defined factors or even by cocktails of chemicals have attracted increased interest because of the broad applications and promises in clinical uses of the induced pluripotent stem cells (iPSCs) (Takahashi and Yamanaka 2006; Hou et al. 2013). Many chromatin proteins are critical for the reprogramming process, such as MBD3 of the NuRD (nucleosome remodeling and deacetylation) complex (Rais et al. 2013), the histone methyltransferases SUV39H and DOT1L (Onder et al. 2012; Soufi et al. 2012), the heterochromatin structural protein HP1 (Sridharan et al. 2013), chromatin remodelers CHD1 and BRG1 (Gaspar-Maia et al. 2009; Singhal et al. 2010), indicating the importance of chromatin in cell reprogramming and cell fate decision. Notably, a *Xenopus* oocyte specific histone H1 variant B4 is required for pluripotency gene reactivation of differentiated ESCs during nuclear reprogramming and an increase of H1 mobility has been observed during reprogramming (Jullien et al. 2010). Although the functions of H1 during transcription factors-induced reprogramming in mammals are unknown, it would be very interesting to investigate the genome localizations of individual H1 variants during reprogramming, which will pave the way for elucidating the functions and dynamics of H1 regulated chromatin compaction in this important process.

Surprisingly, all three H1 variants analyzed in this dissertation are enriched at major satellites, the tandem repeats constituting the pericentric heterochromatin. Such enrichment is specific and is confirmed by qChIP and multiple mapping approaches, suggesting that H1 plays a possible role in maintaining the structure and functions of pericentric heterochromatin. Indeed, H1 enrichment at major satellites leads to the increased nucleosome repeat length compared with bulk chromatin. FISH and qRT-PCR results demonstrate that H1 depletion causes chromocenter clustering and de-repression of major satellites.

The significant binding of H1 at major satellites raises the question why H1 is specifically enriched at these repetitive elements. *In vitro* experiments such as EMSA with reconstituted nucleosomes may elucidate whether major satellites has an intrinsic affinity to H1 proteins or other factors play a role in the binding. On the other hand, the mechanisms through which H1 regulates chromocenter clustering and the level of pericentric transcripts are still unknown. Major satellites harbor consensus binding sites of several transcription factors, within which Pax3 and Pax9 have been reported to redundantly repress pericentric transcription as well as maintain heterochromatin integrity of major satellites (Bulut-Karslioglu et al. 2012). It is possible that H1 may recruit these factors to facilitate the repression of major satellites, thus the relationship between H1 and major satellite bound transcription factors could be pursued as one of the future directions. The timing and level of transcription from major satellites have been shown to be critical for early embryo development (Probst et al. 2010; Santenard et al. 2010; Casanova et al. 2013), in which H1 and higher order chromatin folding plays an important role (Steinbach et al. 1997; Lim et al. 2013; Lin et al. 2013; Perez-Montero et

al. 2013). Exploration of the impact of H1 on pericentric transcripts during early embryo development would provide further understanding on how chromatin compaction affects development.

APPENDIX

PRIMERS USED IN THIS STUDY

Table A.1. List of primers

Name	Number	Sequence	Product Size (bp)	Experimental Purpose
Major	370-MaS-F	GACGACTTGAAAAATGACGAAATC	307	qChIP, qRT-PCR
	371-MaS-R	CATATTCCAGGTCCTTCAGTGTGC		
LINE L1	384-LORF-F	TTTGGGACACAATGAAAGCA	155	qChIP, qRT-PCR
	385-LORF-R	CTGCCGTCTACTCCTCTTGG		
Minor	372-MiS-F	CATGGAAAATGATAAAAACC	162	qChIP, qRT-PCR
	373-MiS-R	CATCTAATATGTTCTACAGTGTGG		
IAP LTR	1421-IAPs-F	GCGCTGACAGCTGTGTTCTA	138	qChIP
	1422-IAPs-R	GGACGTGTCACTCCCTGATT		
Hprt	1427-Hprt-F	CCAAGACGACCGCATGAGAG	101	qChIP
	1428-Hprt-R	CAACGGAGTGATTGCGCATT		
Major_BSF	3180-majorBSF-F	GGAATATGGTAAGAAAATTGAAAATTATGG	360	Bisulfite PCR
	3181-majorBSF-R	CCATATTCCAAATCCTTCAATATACATTTC		
Minor_BSF	3182-minorBSF-F	TAGAATATATTAGATGAGTGAGTTATATTG	208	Bisulfite PCR
	3183-minorBSF-R	ATTATAACTCATTAATATACACTATTCTAC		
Gapdh	89	TTCACCACCATGGAGAAGGC	237	qRT-PCR
	90	GGCATGGACTGTGGTCATGA		

REFERENCES

Ahmed K, Dehghani H, Rugg-Gunn P, Fussner E, Rossant J, Bazett-Jones DP. 2010. Global chromatin architecture reflects pluripotency and lineage commitment in the early mouse embryo. PLoS One 5: e10531.

Alami R, Fan Y, Pack S, Sonbuchner TM, Besse A, Lin Q, Greally JM, Skoultchi AI, Bouhassira EE. 2003. Mammalian linker-histone subtypes differentially affect gene expression in vivo. Proc Natl Acad Sci U S A 100: 5920-5925.

Albig W, Drabent B, Kunz J, Kalff-Suske M, Grzeschik KH, Doenecke D. 1993. All known human H1 histone genes except the H1(0) gene are clustered on chromosome 6. Genomics 16: 649-654.

Albig W, Kardalinou E, Drabent B, Zimmer A, Doenecke D. 1991. Isolation and characterization of two human H1 histone genes within clusters of core histone genes. Genomics 10: 940-948.

Albig W, Meergans T, Doenecke D. 1997. Characterization of the H1.5 gene completes the set of human H1 subtype genes. Gene 184: 141-148.

Allan J, Hartman PG, Crane-Robinson C, Aviles FX. 1980. The structure of histone H1 and its location in chromatin. Nature 288: 675-679.

Bannister AJ, Zegerman P, Partridge JF, Miska EA, Thomas JO, Allshire RC, Kouzarides T. 2001. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature 410: 120-124.

Barra JL, Rhounim L, Rossignol JL, Faugeron G. 2000. Histone H1 is dispensable for methylation-associated gene silencing in Ascobolus immersus and essential for long life span. Mol Cell Biol 20: 61-69.

Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K. 2007. High-resolution profiling of histone methylations in the human genome. Cell 129: 823-837.

Bartolomei MS. 2009. Genomic imprinting: employing and avoiding epigenetic processes. Genes Dev 23: 2124-2133.

Beard C, Li E, Jaenisch R. 1995. Loss of methylation activates Xist in somatic but not in embryonic cells. Genes Dev 9: 2325-2334.

Belancio VP, Roy-Engel AM, Pochampally RR, Deininger P. 2010. Somatic expression of LINE-1 elements in human tissues. Nucleic Acids Res 38: 3909-3922.

Belikov S, Astrand C, Wrange O. 2007. Mechanism of histone H1-stimulated glucocorticoid receptor DNA binding in vivo. Mol Cell Biol 27: 2398-2410.

Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K et al. 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 125: 315-326.

Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, Kellis M, Marra MA, Beaudet AL, Ecker JR et al. 2010. The NIH Roadmap Epigenomics Mapping Consortium. Nat Biotechnol 28: 1045-1048.

Berretta J, Morillon A. 2009. Pervasive transcription constitutes a new level of eukaryotic genome regulation. EMBO Rep 10: 973-982.

Bhan S, May W, Warren SL, Sittman DB. 2008. Global gene expression analysis reveals specific and redundant roles for H1 variants, H1c and H1(0), in gene expression regulation. Gene 414: 10-18.

Birney E Stamatoyannopoulos JA Dutta A Guigo R Gingeras TR Margulies EH Weng Z Snyder M Dermitzakis ET Thurman RE et al. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447: 799-816.

Bock C, Halachev K, Buch J, Lengauer T. 2009. EpiGRAPH: user-friendly software for statistical analysis and prediction of (epi)genomic data. Genome Biol 10: R14.

Bock C, Reither S, Mikeska T, Paulsen M, Walter J, Lengauer T. 2005. BiQ Analyzer: visualization and quality control for DNA methylation data from bisulfite sequencing. Bioinformatics 21: 4067-4068.

Bock C, Tomazou EM, Brinkman AB, Muller F, Simmer F, Gu H, Jager N, Gnirke A, Stunnenberg HG, Meissner A. 2010. Quantitative comparison of genome-wide DNA methylation mapping technologies. Nat Biotechnol 28: 1106-1114.

Bonnefoy E, Bandu MT, Doly J. 1999. Specific binding of high-mobility-group I (HMGI) protein and histone H1 to the upstream AT-rich region of the murine beta interferon promoter: HMGI protein acts as a potential antirepressor of the promoter. Mol Cell Biol 19: 2803-2816.

Bouvet P, Dimitrov S, Wolffe AP. 1994. Specific regulation of Xenopus chromosomal 5S rRNA gene transcription in vivo by histone H1. Genes Dev 8: 1147-1159.

Bouzinba-Segard H, Guais A, Francastel C. 2006. Accumulation of small murine minor satellite transcripts leads to impaired centromeric architecture and function. Proc Natl Acad Sci U S A 103: 8709-8714.

Boyer LA, Plath K, Zeitlinger J, Brambrink T, Medeiros LA, Lee TI, Levine SS, Wernig M, Tajonar A, Ray MK et al. 2006. Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature 441: 349-353.

Bradley A, Evans M, Kaufman MH, Robertson E. 1984. Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. Nature 309: 255-256.

Braunschweig U, Hogan GJ, Pagie L, van Steensel B. 2009. Histone H1 binding is inhibited by histone variant H3.3. EMBO J 28: 3635-3645.

Brero A, Easwaran HP, Nowak D, Grunewald I, Cremer T, Leonhardt H, Cardoso MC. 2005. Methyl CpG-binding proteins induce large-scale chromatin reorganization during terminal differentiation. J Cell Biol 169: 733-743.

Brown DT, Izard T, Misteli T. 2006. Mapping the interaction surface of linker histone H1(0) with the nucleosome of native chromatin in vivo. Nat Struct Mol Biol 13: 250-255.

Bulut-Karslioglu A, Perrera V, Scaranaro M, de la Rosa-Velazquez IA, van de Nobelen S, Shukeir N, Popow J, Gerle B, Opravil S, Pagani M et al. 2012. A transcription factor-based mechanism for mouse heterochromatin formation. Nat Struct Mol Biol 19: 1023-1030.

Campos EI, Reinberg D. 2009. Histones: annotating chromatin. Annu Rev Genet 43: 559-599.

Cao R, Wang L, Wang H, Xia L, Erdjument-Bromage H, Tempst P, Jones RS, Zhang Y. 2002. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science 298: 1039-1043.

Casanova M, Pasternak M, El Marjou F, Le Baccon P, Probst AV, Almouzni G. 2013. Heterochromatin reorganization during early mouse development requires a single-stranded noncoding transcript. Cell Rep 4: 1156-1167.

Chapman GE, Hartman PG, Bradbury EM. 1976. Studies on the role and mode of operation of the very-lysine-rich histone H1 in eukaryote chromatin. The isolation of the globular and non-globular regions of the histone H1 molecule. Eur J Biochem 61: 69-75.

Choo KHA. 1997. *The centromere*. Oxford University Press, Oxford ; New York.

Chow J, Heard E. 2009. X inactivation and the complexities of silencing a sex chromosome. Curr Opin Cell Biol 21: 359-366.

Clausell J, Happel N, Hale TK, Doenecke D, Beato M. 2009. Histone H1 subtypes differentially modulate chromatin condensation without preventing ATP-dependent remodeling by SWI/SNF or NURF. PLoS One 4: e0007243.

Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA et al. 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A 107: 21931-21936.

Creyghton MP, Markoulaki S, Levine SS, Hanna J, Lodato MA, Sha K, Young RA, Jaenisch R, Boyer LA. 2008. H2AZ is enriched at polycomb complex target genes in ES cells and is necessary for lineage commitment. Cell 135: 649-661.

Croston GE, Kerrigan LA, Lira LM, Marshak DR, Kadonaga JT. 1991. Sequence-specific antirepression of histone H1-mediated inhibition of basal RNA polymerase II transcription. Science 251: 643-649.

Daujat S, Zeissler U, Waldmann T, Happel N, Schneider R. 2005. HP1 binds specifically to Lys26-methylated histone H1.4, whereas simultaneous Ser27 phosphorylation blocks HP1 binding. J Biol Chem 280: 38090-38095.

Drabent B, Bode C, Doenecke D. 1993. Structure and expression of the mouse testicular H1 histone gene (H1t). Biochim Biophys Acta 1216: 311-313.

Drabent B, Kardalinou E, Doenecke D. 1991. Structure and expression of the human gene encoding testicular H1 histone (H1t). Gene 103: 263-268.

Drabent B, Saftig P, Bode C, Doenecke D. 2000. Spermatogenesis proceeds normally in mice without linker histone H1t. Histochem Cell Biol 113: 433-442.

Efroni S, Duttagupta R, Cheng J, Dehghani H, Hoeppner DJ, Dash C, Bazett-Jones DP, Le Grice S, McKay RD, Buetow KH et al. 2008. Global transcription in pluripotent embryonic stem cells. Cell Stem Cell 2: 437-447.

Eissenberg JC, Elgin SC. 2000. The HP1 protein family: getting a grip on chromatin. Curr Opin Genet Dev 10: 204-210.

Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M et al. 2011. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature 473: 43-49.

Evans MJ, Kaufman MH. 1981. Establishment in culture of pluripotential cells from mouse embryos. Nature 292: 154-156.

Fan Y, Nikitina T, Morin-Kensicki EM, Zhao J, Magnuson TR, Woodcock CL, Skoultchi AI. 2003. H1 linker histones are essential for mouse development and affect nucleosome spacing in vivo. Mol Cell Biol 23: 4559-4572.

Fan Y, Nikitina T, Zhao J, Fleury TJ, Bhattacharyya R, Bouhassira EE, Stein A, Woodcock CL, Skoultchi AI. 2005. Histone h1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation. Cell 123: 1199-1212.

Fan Y, Sirotkin A, Russell RG, Ayala J, Skoultchi AI. 2001. Individual somatic H1 subtypes are dispensable for mouse development even in mice lacking the H1(0) replacement subtype. Mol Cell Biol 21: 7933-7943.

Fan Y, Skoultchi AI. 2004. Genetic analysis of H1 linker histone subtypes and their functions in mice. Methods Enzymol 377: 85-107.

Fantz DA, Hatfield WR, Horvath G, Kistler MK, Kistler WS. 2001. Mice with a targeted disruption of the H1t gene are fertile and undergo normal changes in structural chromosomal proteins during spermiogenesis. Biol Reprod 64: 425-431.

Freidkin I, Katcoff DJ. 2001. Specific distribution of the Saccharomyces cerevisiae linker histone homolog HHO1p in the chromatin. Nucleic Acids Res 29: 4043-4051.

Gaspar-Maia A, Alajem A, Polesso F, Sridharan R, Mason MJ, Heidersbach A, Ramalho-Santos J, McManus MT, Plath K, Meshorer E et al. 2009. Chd1 regulates open chromatin and pluripotency of embryonic stem cells. Nature 460: 863-868.

George EM, Izard T, Anderson SD, Brown DT. 2010. Nucleosome interaction surface of linker histone H1c is distinct from that of H1(0). J Biol Chem 285: 20891-20896.

Giambra V, Volpi S, Emelyanov AV, Pflugh D, Bothwell AL, Norio P, Fan Y, Ju Z, Skoultchi AI, Hardy RR et al. 2008. Pax5 and linker histone H1 coordinate DNA methylation and histone modifications in the 3' regulatory region of the immunoglobulin heavy chain locus. Mol Cell Biol 28: 6123-6133.

Gilbert N, Allan J. 2001. Distinctive higher-order chromatin structure at mammalian centromeres. Proc Natl Acad Sci U S A 98: 11949-11954.

Gilbert N, Thomson I, Boyle S, Allan J, Ramsahoye B, Bickmore WA. 2007. DNA methylation affects nuclear organization, histone modifications, and linker histone binding but not chromatin compaction. J Cell Biol 177: 401-411.

Guenatri M, Bailly D, Maison C, Almouzni G. 2004. Mouse centric and pericentric satellite repeats form distinct functional heterochromatin. J Cell Biol 166: 493-505.

Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA. 2007. A chromatin landmark and transcription initiation at most promoters in human cells. Cell 130: 77-88.

Gunjan A, Brown DT. 1999. Overproduction of histone H1 variants in vivo increases basal and induced activity of the mouse mammary tumor virus promoter. Nucleic Acids Res 27: 3355-3363.

Hale TK, Contreras A, Morrison AJ, Herrera RE. 2006. Phosphorylation of the linker histone H1 by CDK regulates its binding to HP1alpha. Mol Cell 22: 693-699.

Happel N, Doenecke D. 2009. Histone H1 and its isoforms: contribution to chromatin structure and function. Gene 431: 1-12.

Happel N, Schulze E, Doenecke D. 2005. Characterisation of human histone H1x. Biol Chem 386: 541-551.

Happel N, Warneboldt J, Hanecke K, Haller F, Doenecke D. 2009. H1 subtype expression during cell proliferation and growth arrest. Cell Cycle 8: 2226-2232.

Harnicarova Horakova A, Galiova G, Legartova S, Kozubek S, Matula P, Bartova E. 2010. Chromocentre integrity and epigenetic marks. J Struct Biol 169: 124-133.

Hashimoto H, Takami Y, Sonoda E, Iwasaki T, Iwano H, Tachibana M, Takeda S, Nakayama T, Kimura H, Shinkai Y. 2010. Histone H1 null vertebrate cells exhibit altered nucleosome architecture. Nucleic Acids Res 38: 3533-3545.

Hecht A, Strahl-Bolsinger S, Grunstein M. 1996. Spreading of transcriptional repressor SIR3 from telomeric heterochromatin. Nature 383: 92-96.

Hellauer K, Sirard E, Turcotte B. 2001. Decreased expression of specific genes in yeast cells lacking histone H1. J Biol Chem 276: 13587-13592.

Hendzel MJ, Lever MA, Crawford E, Th'ng JP. 2004. The C-terminal domain is the primary determinant of histone H1 binding to chromatin in vivo. J Biol Chem 279: 20028-20034.

Hockemeyer D, Wang H, Kiani S, Lai CS, Gao Q, Cassady JP, Cost GJ, Zhang L, Santiago Y, Miller JC et al. 2011. Genetic engineering of human pluripotent cells using TALE nucleases. Nat Biotechnol 29: 731-734.

Hou P, Li Y, Zhang X, Liu C, Guan J, Li H, Zhao T, Ye J, Yang W, Liu K et al. 2013. Pluripotent stem cells induced from mouse somatic cells by small-molecule compounds. Science 341: 651-654.

Howell CY, Bestor TH, Ding F, Latham KE, Mertineit C, Trasler JM, Chaillet JR. 2001. Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene. Cell 104: 829-838.

Hsu TC, Cooper JE, Mace ML, Jr., Brinkley BR. 1971. Arrangement of centromeres in mouse cells. Chromosoma 34: 73-87.

Izaurralde E, Kas E, Laemmli UK. 1989. Highly preferential nucleation of histone H1 assembly on scaffold-associated regions. J Mol Biol 210: 573-585.

Izzo A, Kamieniarz-Gdula K, Ramirez F, Noureen N, Kind J, Manke T, van Steensel B, Schneider R. 2013. The genomic landscape of the somatic linker histone subtypes H1.1 to H1.5 in human cells. Cell Rep 3: 2142-2154.

Izzo A, Kamieniarz K, Schneider R. 2008. The histone H1 family: specific members, specific functions? Biol Chem 389: 333-343.

Jackson-Grusby L, Beard C, Possemato R, Tudor M, Fambrough D, Csankovszki G, Dausman J, Lee P, Wilson C, Lander E et al. 2001. Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. Nat Genet 27: 31-39.

Jedrusik MA, Schulze E. 2001. A single histone H1 isoform (H1.1) is essential for chromatin silencing and germline development in Caenorhabditis elegans. Development 128: 1069-1080.

Jeong S, Liang G, Sharma S, Lin JC, Choi SH, Han H, Yoo CB, Egger G, Yang AS, Jones PA. 2009. Selective anchoring of DNA methyltransferases 3A and 3B to nucleosomes containing methylated DNA. Mol Cell Biol 29: 5366-5376.

Jones DO, Cowell IG, Singh PB. 2000. Mammalian chromodomain proteins: their role in genome organisation and expression. Bioessays 22: 124-137.

Jullien J, Astrand C, Halley-Stott RP, Garrett N, Gurdon JB. 2010. Characterization of somatic cell nuclear reprogramming by oocytes in which a linker histone is

required for pluripotency gene reactivation. Proc Natl Acad Sci U S A 107: 5483-5488.

Jurka J. 2000. Repbase update: a database and an electronic journal of repetitive elements. Trends Genet 16: 418-420.

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. 2005. Repbase Update, a database of eukaryotic repetitive elements. Cytogenet Genome Res 110: 462-467.

Jurkowska RZ, Jurkowski TP, Jeltsch A. 2011. Structure and function of mammalian DNA methyltransferases. Chembiochem 12: 206-222.

Kandolf H. 1994. The H1A histone variant is an in vivo repressor of oocyte-type 5S gene transcription in Xenopus laevis embryos. Proc Natl Acad Sci U S A 91: 7257-7261.

Kas E, Izaurralde E, Laemmli UK. 1989. Specific inhibition of DNA binding to nuclear scaffolds and histone H1 by distamycin. The role of oligo(dA).oligo(dT) tracts. J Mol Biol 210: 587-599.

Kashiwagi K, Nimura K, Ura K, Kaneda Y. 2011. DNA methyltransferase 3b preferentially associates with condensed chromatin. Nucleic Acids Res 39: 874-888.

Kidder BL, Yang J, Palmer S. 2008. Stat3 and c-Myc genome-wide promoter occupancy in embryonic stem cells. PLoS One 3: e3932.

Kouzarides T. 2007. Chromatin modifications and their function. Cell 128: 693-705.

Krishnakumar R, Gamble MJ, Frizzell KM, Berrocal JG, Kininis M, Kraus WL. 2008. Reciprocal binding of PARP-1 and histone H1 at promoters specifies transcriptional outcomes. Science 319: 819-821.

Lachner M, O'Carroll D, Rea S, Mechtler K, Jenuwein T. 2001. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature 410: 116-120.

Lajugie J, Bouhassira EE. 2011. GenPlay, a multipurpose genome analyzer and browser. Bioinformatics 27: 1889-1893.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10: R25.

Laybourn PJ, Kadonaga JT. 1991. Role of nucleosomal cores and histone H1 in regulation of transcription by RNA polymerase II. Science 254: 238-245.

Lee H, Habas R, Abate-Shen C. 2004. MSX1 cooperates with histone H1b for inhibition of transcription and myogenesis. Science 304: 1675-1678.

Lehnertz B, Ueda Y, Derijck AA, Braunschweig U, Perez-Burgos L, Kubicek S, Chen T, Li E, Jenuwein T, Peters AH. 2003. Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. Curr Biol 13: 1192-1200.

Lennox RW, Cohen LH. 1983. The histone H1 complements of dividing and nondividing cells of the mouse. J Biol Chem 258: 262-268.

Lever MA, Th'ng JP, Sun X, Hendzel MJ. 2000. Rapid exchange of histone H1.1 on chromatin in living human cells. Nature 408: 873-876.

Lewis JD, Meehan RR, Henzel WJ, Maurer-Fogy I, Jeppesen P, Klein F, Bird A. 1992. Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. Cell 69: 905-914.

Li E, Beard C, Jaenisch R. 1993. Role for DNA methylation in genomic imprinting. Nature 366: 362-365.

Li JY, Patterson M, Mikkola HK, Lowry WE, Kurdistani SK. 2012. Dynamic distribution of linker histone h1.5 in cellular differentiation. PLoS Genet 8: e1002879.

Lim CY, Reversade B, Knowles BB, Solter D. 2013. Optimal histone H3 to linker histone H1 chromatin ratio is vital for mesodermal competence in Xenopus. Development 140: 853-860.

Lin CJ, Conti M, Ramalho-Santos M. 2013. Histone variant H3.3 maintains a decondensed chromatin state essential for mouse preimplantation development. Development 140: 3624-3634.

Lin Q, Inselman A, Han X, Xu H, Zhang W, Handel MA, Skoultchi AI. 2004. Reductions in linker histone levels are tolerated in developing spermatocytes but cause changes in specific gene expression. J Biol Chem 279: 23525-23535.

Lin Q, Sirotkin A, Skoultchi AI. 2000. Normal spermatogenesis in mice lacking the testis-specific linker histone H1t. Mol Cell Biol 20: 2122-2128.

Linxweller W, Horz W. 1985. Reconstitution experiments show that sequence-specific histone-DNA interactions are the basis for nucleosome phasing on mouse satellite DNA. Cell 42: 281-290.

Lu R, Markowetz F, Unwin RD, Leek JT, Airoldi EM, MacArthur BD, Lachmann A, Rozov R, Ma'ayan A, Boyer LA et al. 2009a. Systems-level dynamic analyses of fate change in murine embryonic stem cells. Nature 462: 358-362.

Lu X, Wontakal SN, Emelyanov AV, Morcillo P, Konev AY, Fyodorov DV, Skoultchi AI. 2009b. Linker histone H1 is essential for Drosophila development, the establishment of pericentric heterochromatin, and a normal polytene chromosome structure. Genes Dev 23: 452-465.

Lu X, Wontakal SN, Kavi H, Kim BJ, Guzzardo PM, Emelyanov AV, Xu N, Hannon GJ, Zavadil J, Fyodorov DV et al. 2013. Drosophila H1 regulates the genetic activity of heterochromatin by recruitment of Su(var)3-9. Science 340: 78-81.

Lukiw WJ, Kruck TP, McLachlan DR. 1989. Linker histone-DNA complexes: enhanced stability in the presence of aluminum lactate and implications for Alzheimer's disease. FEBS Lett 253: 59-62.

Maclean JA, Bettegowda A, Kim BJ, Lou CH, Yang SM, Bhardwaj A, Shanker S, Hu Z, Fan Y, Eckardt S et al. 2011. The rhox homeobox gene cluster is imprinted and selectively targeted for regulation by histone h1 and DNA methylation. Mol Cell Biol 31: 1275-1287.

Maison C, Bailly D, Peters AH, Quivy JP, Roche D, Taddei A, Lachner M, Jenuwein T, Almouzni G. 2002. Higher-order structure in pericentric heterochromatin involves

a distinct pattern of histone modification and an RNA component. Nat Genet 30: 329-334.

Maison C, Bailly D, Roche D, Montes de Oca R, Probst AV, Vassias I, Dingli F, Lombard B, Loew D, Quivy JP et al. 2011. SUMOylation promotes de novo targeting of HP1alpha to pericentric heterochromatin. Nat Genet 43: 220-227.

Mansour SL, Thomas KR, Capecchi MR. 1988. Disruption of the proto-oncogene int-2 in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. Nature 336: 348-352.

Margueron R, Reinberg D. 2011. The Polycomb complex PRC2 and its mark in life. Nature 469: 343-349.

Martens JH, O'Sullivan RJ, Braunschweig U, Opravil S, Radolf M, Steinlein P, Jenuwein T. 2005. The profile of repeat-associated histone lysine methylation states in the mouse epigenome. EMBO J 24: 800-812.

Martianov I, Brancorsini S, Catena R, Gansmuller A, Kotaja N, Parvinen M, Sassone-Corsi P, Davidson I. 2005. Polar nuclear localization of H1T2, a histone H1 variant, required for spermatid elongation and DNA condensation during spermiogenesis. Proc Natl Acad Sci U S A 102: 2808-2813.

Martin C, Cao R, Zhang Y. 2006. Substrate preferences of the EZH2 histone methyltransferase complex. J Biol Chem 281: 8365-8370.

Medrzycki M, Zhang Y, Cao K, Fan Y. 2012a. Expression Analysis of Mammalian Linker-histone Subtypes. J Vis Exp 61: e3577, DOI :3510.3791/3577.

Medrzycki M, Zhang Y, McDonald JF, Fan Y. 2012b. Profiling of linker histone variants in ovarian cancer. Front Biosci 17: 396-406.

Meergans T, Albig W, Doenecke D. 1997. Varied expression patterns of human H1 histone genes in different cell lines. DNA Cell Biol 16: 1041-1049.

Meissner A. 2010. Epigenetic modifications in pluripotent and differentiated cells. Nat Biotechnol 28: 1079-1088.

Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB et al. 2008. Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature 454: 766-770.

Meshorer E, Yellajoshula D, George E, Scambler PJ, Brown DT, Misteli T. 2006. Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. Dev Cell 10: 105-116.

Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP et al. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 448: 553-560.

Miller JC, Holmes MC, Wang J, Guschin DY, Lee YL, Rupniewski I, Beausejour CM, Waite AJ, Wang NS, Kim KA et al. 2007. An improved zinc-finger nuclease architecture for highly specific genome editing. Nat Biotechnol 25: 778-785.

Mirkovitch J, Mirault ME, Laemmli UK. 1984. Organization of the higher-order chromatin loop: specific DNA attachment sites on nuclear scaffold. Cell 39: 223-232.

Misteli T, Gunjan A, Hock R, Bustin M, Brown DT. 2000. Dynamic binding of histone H1 to chromatin in living cells. Nature 408: 877-881.

Ni JQ, Liu LP, Hess D, Rietdorf J, Sun FL. 2006. Drosophila ribosomal proteins are associated with linker histone H1 and suppress gene transcription. Genes Dev 20: 1959-1973.

Nielsen AL, Oulad-Abdelghani M, Ortiz JA, Remboutsika E, Chambon P, Losson R. 2001. Heterochromatin formation in mammalian cells: interaction between histones and HP1 proteins. Mol Cell 7: 729-739.

Nishiyama M, Oshikawa K, Tsukada Y, Nakagawa T, Iemura S, Natsume T, Fan Y, Kikuchi A, Skoultchi AI, Nakayama KI. 2009. CHD8 suppresses p53-mediated apoptosis through histone H1 recruitment during early embryogenesis. Nat Cell Biol 11: 172-182.

Nishiyama M, Skoultchi AI, Nakayama KI. 2012. Histone H1 recruitment by CHD8 is essential for suppression of the Wnt-beta-catenin signaling pathway. Mol Cell Biol 32: 501-512.

Noll M, Kornberg RD. 1977. Action of micrococcal nuclease on chromatin and the location of histone H1. J Mol Biol 109: 393-404.

Oberg C, Izzo A, Schneider R, Wrange O, Belikov S. 2012. Linker histone subtypes differ in their effect on nucleosomal spacing in vivo. J Mol Biol 419: 183-197.

Olins AL, Olins DE. 1974. Spheroid chromatin units (v bodies). Science 183: 330-332.

Onder TT, Kara N, Cherry A, Sinha AU, Zhu N, Bernt KM, Cahan P, Marcarci BO, Unternaehrer J, Gupta PB et al. 2012. Chromatin-modifying enzymes as modulators of reprogramming. Nature 483: 598-602.

Orkin SH, Hochedlinger K. 2011. Chromatin connections to pluripotency and cellular reprogramming. Cell 145: 835-850.

Oudet P, Gross-Bellard M, Chambon P. 1975. Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. Cell 4: 281-300.

Packer AI, Manova K, Bachvarova RF. 1993. A discrete LINE-1 transcript in mouse blastocysts. Dev Biol 157: 281-283.

Panetta G, Buttinelli M, Flaus A, Richmond TJ, Rhodes D. 1998. Differential nucleosome positioning on Xenopus oocyte and somatic 5 S RNA genes determines both TFIIIA and H1 binding: a mechanism for selective H1 repression. J Mol Biol 282: 683-697.

Papait R, Pistore C, Grazini U, Babbio F, Cogliati S, Pecoraro D, Brino L, Morand AL, Dechampesme AM, Spada F et al. 2008. The PHD domain of Np95 (mUHRF1) is involved in large-scale reorganization of pericentromeric heterochromatin. Mol Biol Cell 19: 3554-3563.

Park PJ. 2009. ChIP-seq: advantages and challenges of a maturing technology. Nat Rev Genet 10: 669-680.

Perez-Montero S, Carbonell A, Moran T, Vaquero A, Azorin F. 2013. The Embryonic Linker Histone H1 Variant of Drosophila, dBigH1, Regulates Zygotic Genome Activation. Dev Cell 26: 578-590.

Peters AH, O'Carroll D, Scherthan H, Mechtler K, Sauer S, Schofer C, Weipoltshammer K, Pagani M, Lachner M, Kohlmaier A et al. 2001. Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. Cell 107: 323-337.

Probst AV, Almouzni G. 2008. Pericentric heterochromatin: dynamic organization during early development in mammals. Differentiation 76: 15-23.

Probst AV, Almouzni G. 2011. Heterochromatin establishment in the context of genome-wide epigenetic reprogramming. Trends Genet 27: 177-185.

Probst AV, Okamoto I, Casanova M, El Marjou F, Le Baccon P, Almouzni G. 2010. A strand-specific burst in transcription of pericentric satellites is required for chromocenter formation and early mouse development. Dev Cell 19: 625-638.

Rabini S, Franke K, Saftig P, Bode C, Doenecke D, Drabent B. 2000. Spermatogenesis in mice is not affected by histone H1.1 deficiency. Exp Cell Res 255: 114-124.

Rais Y, Zviran A, Geula S, Gafni O, Chomsky E, Viukov S, Mansour AA, Caspi I, Krupalnik V, Zerbib M et al. 2013. Deterministic direct reprogramming of somatic cells to pluripotency. Nature 502: 65-70.

Robertson KD, Wolffe AP. 2000. DNA methylation in health and disease. Nat Rev Genet 1: 11-19.

Routh A, Sandin S, Rhodes D. 2008. Nucleosome repeat length and linker histone stoichiometry determine chromatin fiber structure. Proc Natl Acad Sci U S A 105: 8872-8877.

Rudert F, Bronner S, Garnier JM, Dolle P. 1995. Transcripts from opposite strands of gamma satellite DNA are differentially expressed during mouse development. Mamm Genome 6: 76-83.

Sancho M, Diani E, Beato M, Jordan A. 2008. Depletion of human histone H1 variants uncovers specific roles in gene expression and cell growth. PLoS Genet 4: e1000227.

Santenard A, Ziegler-Birling C, Koch M, Tora L, Bannister AJ, Torres-Padilla ME. 2010. Heterochromatin formation in the mouse embryo requires critical residues of the histone variant H3.3. Nat Cell Biol 12: 853-862.

Shahbazian MD, Grunstein M. 2007. Functions of site-specific histone acetylation and deacetylation. Annu Rev Biochem 76: 75-100.

Sharma S, De Carvalho DD, Jeong S, Jones PA, Liang G. 2011. Nucleosomes containing methylated DNA stabilize DNA methyltransferases 3A/3B and ensure faithful epigenetic inheritance. PLoS Genet 7: e1001286.

Shen X, Gorovsky MA. 1996. Linker histone H1 regulates specific gene expression but not global transcription in vivo. Cell 86: 475-483.

Shen X, Liu Y, Hsu YJ, Fujiwara Y, Kim J, Mao X, Yuan GC, Orkin SH. 2008. EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. Mol Cell 32: 491-502.

Shendure J, Ji H. 2008. Next-generation DNA sequencing. Nat Biotechnol 26: 1135-1145.

Shimamura A, Sapp M, Rodriguez-Campos A, Worcel A. 1989. Histone H1 represses transcription from minichromosomes assembled in vitro. Mol Cell Biol 9: 5573-5584.

Shin H, Liu T, Manrai AK, Liu XS. 2009. CEAS: cis-regulatory element annotation system. Bioinformatics 25: 2605-2606.

Singhal N, Graumann J, Wu G, Arauzo-Bravo MJ, Han DW, Greber B, Gentile L, Mann M, Scholer HR. 2010. Chromatin-Remodeling Components of the BAF Complex Facilitate Reprogramming. Cell 141: 943-955.

Sirotkin AM, Edelmann W, Cheng G, Klein-Szanto A, Kucherlapati R, Skoultchi AI. 1995. Mice develop normally without the H1(0) linker histone. Proc Natl Acad Sci U S A 92: 6434-6438.

Soufi A, Donahue G, Zaret KS. 2012. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. Cell 151: 994-1004.

Sridharan R, Gonzales-Cope M, Chronis C, Bonora G, McKee R, Huang C, Patel S, Lopez D, Mishra N, Pellegrini M et al. 2013. Proteomic and genomic approaches reveal critical functions of H3K9 methylation and heterochromatin protein-1gamma in reprogramming to pluripotency. Nat Cell Biol 15: 872-882.

Stasevich TJ, Mueller F, Brown DT, McNally JG. 2010. Dissecting the binding mechanism of the linker histone in live cells: an integrated FRAP analysis. EMBO J 29: 1225-1234.

Steinbach OC, Wolffe AP, Rupp RA. 1997. Somatic linker histones cause loss of mesodermal competence in Xenopus. Nature 389: 395-399.

Suzuki MM, Bird A. 2008. DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet 9: 465-476.

Syed SH, Goutte-Gattat D, Becker N, Meyer S, Shukla MS, Hayes JJ, Everaers R, Angelov D, Bednar J, Dimitrov S. 2010. Single-base resolution mapping of H1-nucleosome interactions and 3D organization of the nucleosome. Proc Natl Acad Sci U S A 107: 9620-9625.

Takahashi K, Yamanaka S. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126: 663-676.

Takami Y, Nakayama T. 1997. A single copy of linker H1 genes is enough for proliferation of the DT40 chicken B cell line, and linker H1 variants participate in regulation of gene expression. Genes Cells 2: 711-723.

Takami Y, Nishi R, Nakayama T. 2000. Histone H1 variants play individual roles in transcription regulation in the DT40 chicken B cell line. Biochem Biophys Res Commun 268: 501-508.

Tanaka H, Iguchi N, Isotani A, Kitamura K, Toyama Y, Matsuoka Y, Onishi M, Masai K, Maekawa M, Toshimori K et al. 2005. HANP1/H1T2, a novel histone H1-like protein involved in nuclear formation and sperm fertility. Mol Cell Biol 25: 7107-7119.

Tanaka M, Hennebold JD, Macfarlane J, Adashi EY. 2001. A mammalian oocyte-specific linker histone gene H1oo: homology with the genes for the oocyte-specific cleavage stage histone (cs-H1) of sea urchin and the B4/H1M histone of the frog. Development 128: 655-664.

Th'ng JP, Sung R, Ye M, Hendzel MJ. 2005. H1 family histones in the nucleus. Control of binding and localization by the C-terminal domain. J Biol Chem 280: 27809-27814.

Thoma F, Koller T. 1977. Influence of histone H1 on chromatin structure. Cell 12: 101-107.

Thoma F, Koller T, Klug A. 1979. Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. J Cell Biol 83: 403-427.

Thomas JO, Rees C, Finch JT. 1992. Cooperative binding of the globular domains of histones H1 and H5 to DNA. Nucleic Acids Res 20: 187-194.

Tillo D, Hughes TR. 2009. G+C content dominates intrinsic nucleosome occupancy. BMC Bioinformatics 10: 442.

Trelogan SA, Martin SL. 1995. Tightly regulated, developmentally specific expression of the first open reading frame from LINE-1 during mouse embryogenesis. Proc Natl Acad Sci U S A 92: 1520-1524.

Trojer P, Reinberg D. 2007. Facultative heterochromatin: is there a distinctive molecular signature? Mol Cell 28: 1-13.

van Holde KE. 1989. *Chromatin*. New York: Springer-Verlag.

Vaquero A, Scher M, Lee D, Erdjument-Bromage H, Tempst P, Reinberg D. 2004. Human SirT1 interacts with histone H1 and promotes formation of facultative heterochromatin. Mol Cell 16: 93-105.

Vyas P, Brown DT. 2012. N- and C-terminal domains determine differential nucleosomal binding geometry and affinity of linker histone isotypes H1(0) and H1c. J Biol Chem 287: 11778-11787.

Walsh CP, Chaillet JR, Bestor TH. 1998. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. Nat Genet 20: 116-117.

Wang H, Yang H, Shivalila CS, Dawlaty MM, Cheng AW, Zhang F, Jaenisch R. 2013. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell 153: 910-918.

Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, Cuddapah S, Cui K, Roh TY, Peng W, Zhang MQ et al. 2008. Combinatorial patterns of histone acetylations and methylations in the human genome. Nat Genet 40: 897-903.

Wang ZF, Sirotkin AM, Buchold GM, Skoultchi AI, Marzluff WF. 1997. The mouse histone H1 genes: gene organization and differential regulation. J Mol Biol 271: 124-138.

Warneboldt J, Haller F, Horstmann O, Danner BC, Fuzesi L, Doenecke D, Happel N. 2008. Histone H1x is highly expressed in human neuroendocrine cells and tumours. BMC Cancer 8: 388.

Wierzbicki AT, Jerzmanowski A. 2005. Suppression of histone H1 genes in Arabidopsis results in heritable developmental defects and stochastic changes in DNA methylation. Genetics 169: 997-1008.

Wolffe AP. 1998. *Chromatin: Structure and Function*. Academic Press, San Diego, CA, San Diego, CA.

Woodcock CL, Skoultchi AI, Fan Y. 2006. Role of linker histone in chromatin structure and function: H1 stoichiometry and nucleosome repeat length. Chromosome Res 14: 17-25.

Yamagata K, Yamazaki T, Miki H, Ogonuki N, Inoue K, Ogura A, Baba T. 2007. Centromeric DNA hypomethylation as an epigenetic signature discriminates between germ and somatic cell lineages. Dev Biol 312: 419-426.

Yan W, Ma L, Burns KH, Matzuk MM. 2003. HILS1 is a spermatid-specific linker histone H1-like protein implicated in chromatin remodeling during mammalian spermiogenesis. Proc Natl Acad Sci U S A 100: 10546-10551.

Yang SM, Kim BJ, Norwood Toro L, Skoultchi AI. 2013. H1 linker histone promotes epigenetic silencing by regulating both DNA methylation and histone H3 methylation. Proc Natl Acad Sci U S A 110: 1708-1713.

Young RA. 2011. Control of the embryonic stem cell state. Cell 144: 940-954.

Zang C, Schones DE, Zeng C, Cui K, Zhao K, Peng W. 2009. A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. Bioinformatics 25: 1952-1958.

Zhang XY, Horz W. 1984. Nucleosomes are positioned on mouse satellite DNA in multiple highly specific frames that are correlated with a diverged subrepeat of nine base-pairs. J Mol Biol 176: 105-129.

Zhang Y, Cooke M, Panjwani S, Cao K, Krauth B, Ho PY, Medrzycki M, Berhe DT, Pan C, McDevitt TC et al. 2012a. Histone h1 depletion impairs embryonic stem cell differentiation. PLoS Genet 8: e1002691.

Zhang Y, Liu Z, Medrzycki M, Cao K, Fan Y. 2012b. Reduction of Hox gene expression by histone H1 depletion. PLoS One 7: e38829.

Zhou J, Saleque S, Ermakova O, Sepulveda MA, Yang Q, Eckhardt LA, Schildkraut CL, Birshtein BK. 2005. Changes in replication, nuclear location, and expression of the Igh locus after fusion of a pre-B cell line with a T cell line. J Immunol 175: 2317-2320.

Zlatanova J, Doenecke D. 1994. Histone H1 zero: a major player in cell differentiation? FASEB J 8: 1260-1268.